2002-06

How agent based models can be utilized to explore and exploit non-linearity and intangibles inherent in guerrilla warfare

Ipekci, Arif Ilker

Monterey, Calif. Naval Postgraduate School

http://hdl.handle.net/10945/5907

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

HOW AGENT BASED MODELS CAN BE UTILIZED TO EXPLORE AND EXPLOIT NON-LINEARITY AND INTANGIBLES INHERENT IN GUERRILLA WARFARE

by

Arif Ilker Ipekci

June 2002

Thesis Advisor:	Thomas Lucas
Second Reader:	Saverio Manago

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2002	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE: How Agent Based Models Can Be Utilized To Explore and Exploit Non-linearity and Intangibles Inherent in Guerrilla Warfare

5. FUNDING NUMBERS

6. AUTHOR(S) Arif Ilker Ipekci

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approve for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

Since the end of WWII, a host of groups and states have pursued their interests in the Low Intensity Conflict (LIC) environment. One of the characteristics of LIC is that it is executed mostly by the rules of asymmetric war or guerrilla warfare. This thesis utilizes the recently developed agent-based model Map Aware Non-Uniform Automata (MANA) to explore non-linearity and intangibles inherent in guerrilla warfare. An infiltration scenario is developed based on the author's experiences fighting guerrillas in the mountains of Southeast Turkey. To simultaneously investigate the effects of as many as 22 input variables, recently developed Near Orthogonal Latin Hypercube Designs and Fractional Factorial Designs are used. Utilizing a personal computer and the computational capabilities of supercomputers run by Mitre for the Marine Corps Combat Development Center (MCCDC), approximately 200,000 MANA runs were completed. Several statistical models are developed and compared using a variety of diverse statistical techniques, including Cluster Analysis, Neural Networks, Regression Trees, Linear Regression, and Bayesian Networks. The results of the analysis suggest that the outcome of an infiltration scenario is heavily dependent on the Red agent parameters. The analysis also reveals the Red Stealth parameter as the most important factor in predicting the MOEs.

14. SUBJECT TERMS Agent Based Models, Guerrilla Warfare, Exploring non-linearity and intangibles, Terrorism, Infiltration, Complex Adaptive Systems, Latin Hypercube Designs, Bayesian Networks, Regression Trees, Neural Networks, Cluster Analysis, Data mining applications, Low Intensity Conflict, MANA

15. NUMBER OF PAGES
178

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UL

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited.

HOW AGENT BASED MODELS CAN BE UTILIZED TO EXPLORE AND EXPLOIT NON-LINEARITY AND INTANGIBLES INHERENT IN GUERRILLA WARFARE

Arif Ilker Ipekci
First Lieutenant, Turkish Army
B.S., Turkish Army Academy, 1996

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2002

Author: Arif Ilker Ipekci

Approved by: Thomas Lucas
Thesis Advisor

Saverio Manago
Second Reader

James N. Eagle
Chairman, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Since the end of WWII, a host of groups and states have pursued their interests in the Low Intensity Conflict (LIC) environment. One of the characteristics of LIC is that it is executed mostly by the rules of asymmetric war or guerrilla warfare. This thesis utilizes the recently developed agent-based model Map Aware Non-Uniform Automata (MANA) to explore non-linearity and intangibles inherent in guerrilla warfare. An infiltration scenario is developed based on the author's experiences fighting guerrillas in the mountains of Southeast Turkey. To simultaneously investigate the effects of as many as 22 input variables, recently developed Near Orthogonal Latin Hypercube Designs and Fractional Factorial Designs are used. Utilizing a personal computer and the computational capabilities of supercomputers run by Mitre for the Marine Corps Combat Development Center (MCCDC), approximately 200,000 MANA runs were completed. Several statistical models are developed and compared using a variety of diverse statistical techniques, including Cluster Analysis, Neural Networks, Regression Trees, Linear Regression, and Bayesian Networks. The results of the analysis suggest that the outcome of an infiltration scenario is heavily dependent on the Red agent parameters. The analysis also reveals the Red Stealth parameter as the most important factor in predicting the MOEs.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF SYMBOLS, ACRONYMS AND/OR ABBREVIATIONS

CCF	The Chinese Communist Forces
MCP	The Malayan Communist Party
MPAJA	The Malayan People's Anti-Japanese Army
TNKU	Northern Borneo National Army
NCO	Non-Commissioned Officer
PKK	Kurdish Workers Party
OOTW	Operations Other Than War
GKK	Temporary Village Guards
LOC	Lines of Communication
MANA	Map Aware Non-Uniform Automata
MCCDC	Marine Corps Combat Development Center
RPG	Anti-Tank Grenade Launcher
ISAAC	Irreducible Semi-Autonomous Adaptive Combat
LIC	Low Intensity Conflict
OR	Operations Research
LE	Lanchester Equations
ABM	Agent-Based Model
EBM	Equation-Based Model
CAS	Complex Adaptive System
MCCDC	Marine Corps Combat Development Center
DTA	Defense Technology Agency, New Zealand
SSKP	Single Shot Kill Probability
CA	Cellular Automaton
ACV	Armored Combat Vehicle
NCO	Non-Commissioned Officer
SA	Situational Awareness
HQ	Headquarters
GUI	Graphical User Interface
AI	Artificial Intelligence
MOE	Measure of Effectiveness
LHS	Latin Hypercube Sampling
LHD	Latin Hypercube Design
MART	Multi Additive Regression Trees
ANOVA	Analysis of Variance
GAM	Generalized Additive Model
2-D	Two-Dimensional
3-D	Three-Dimensional
BBN	Bayesian Belief Network
AAE	Average Absolute Error
RSME	Root Mean Squared Error
NPS	Naval Postgraduate School

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to thank Professor Tom W. Lucas for his combat modeling insight, persistence, and guidance during this effort, from our initial discussions to the final product. I would also like to thank Lieutenant Colonel Saverio Manago for his military perspectives, statistical insights and military resources.

I would like to extend a special thanks to Lieutenant Colonel Tom Cioppa for his state of the art Near Orthogonal Latin Hypercube designs. Additionally, I would like to thank Dr. Gary Horne, John Kresho and Sarah Johnson at Marines Corps Combat Development Command for their efforts and support in the data collection phase of the thesis.

Finally, I would like to thank my editor, Beth Summe, for her assistance in editing the chapters of this thesis.

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Since the end of WWII, a host of groups and states have pursued their interests in the Low Intensity Conflict (LIC) environment [Ref.11]. One of the characteristics of LIC is that it is executed mostly by the rules of asymmetric war or guerrilla warfare. Guerrilla warfare is a set of military tactics utilized by a group within a state or a local population in order to oppose the government or foreign occupying forces [Ref.5]. In essence, guerrilla warfare determines the shape of the battlefield in the Low Intensity Conflict environment while remaining a potent weapon for which armed forces must increasingly be prepared. One of the best examples of guerrilla warfare is in Turkey—where conventional Turkish Forces, with more than 100,000 soldiers, have been continuously fighting against approximately 10,000 PKK terrorists for more than 15 years. During this conflict, more than 30,000 people have been killed, including about 4,000 Turkish soldiers.

Recently, many analysts are realizing that conventional Operations Research tools based on precise mathematical equations and a detailed physical description of combat cannot provide a realistic picture of the complex and dynamic situations in which many military operations, particularly guerrilla-type operations, are conducted. With the advent of complexity theory and its application to warfare studies, some researchers are viewing warfare as a complex adaptive system (CAS), which adapts, evolves and co-evolves with its environment [Ref.6]. In this sense, a guerrilla force also behaves like a complex system, not only interacting with its environment but also dynamically (and often nonlinear) interacting between components of the system, namely the interactions between different levels of commanders and the commanders with the enemy. In summary, the focus is on the emergent patterns of the *whole* rather than the individual *part*s.

Agent-based models (ABM) offer an opportunity to analyze the above-mentioned complex problems by concentrating on the behaviors of and interactions between the participating entities instead of the performance of specific weapons or sensors [Ref.4].

This is particularly important in guerrilla warfare where one side usually dominates with respect to firepower. In other words, we shift our attention from analyzing the performance of pieces of equipment to how different modes of operation and human traits may alter the outcome of a combat or peacekeeping operation. Attention is focused on how a guerrilla force utilizes information and acts upon it, and what the consequences are on the success of a mission.

The use of intelligent agents to study the emergent behaviors of complex systems and operations, such as the roles played by training, aggressiveness or the effectiveness of command-and-control in combat performance, has attracted much attention in the military operations research (OR) community in recent years [Ref.4]. Notable examples are models like the Irreducible Semi-Autonomous Adaptive Combat (ISAAC) [Ref.6] initiated by the US Marine Corps Combat Development Command (MCCDC) and the Map Aware Non-uniform Automata (MANA) [Ref.1] developed by the Defense Technology Agency (DTA), New Zealand.

This thesis utilizes the recently developed agent-based model Map Aware Non-Uniform Automata (MANA) to explore non-linearity and intangibles inherent in guerrilla warfare.

An infiltration scenario is developed based on the author's experiences fighting guerrillas in the mountains of Southeast Turkey. In September 1999, I was a platoon commander of a battalion stationed in one of the rural areas in the Southeast of Turkey. The platoon received a mission order from the battalion commander to form a tank-infantry platoon and to move immediately to the top of a nearby hill. The platoon's task was to take up a position on the hill and protect the area against any terrorist activities. This location was very critical for both sides, because terrorists were using this territory as a passage from northern Iraq to Turkey. From this peak, the platoon could easily monitor and interdict the terrorists. Consequently, we expected the terrorists to attempt sneaking into our base to damage assets, cause some casualties, and demolish morale.

In this scenario, the Blue force, which represents the friendly side, is composed of two tanks, two ACVs (Armored Combat Vehicle), and 11 infantry. The Red force, which

represents terrorists, is composed of 11 guerrillas equipped with light infantry weapons. Two of these terrorists form a reconnaissance team. The remaining nine terrorists split into two infiltration teams. The screenshot below is the base-line scenario; it is best viewed in color.

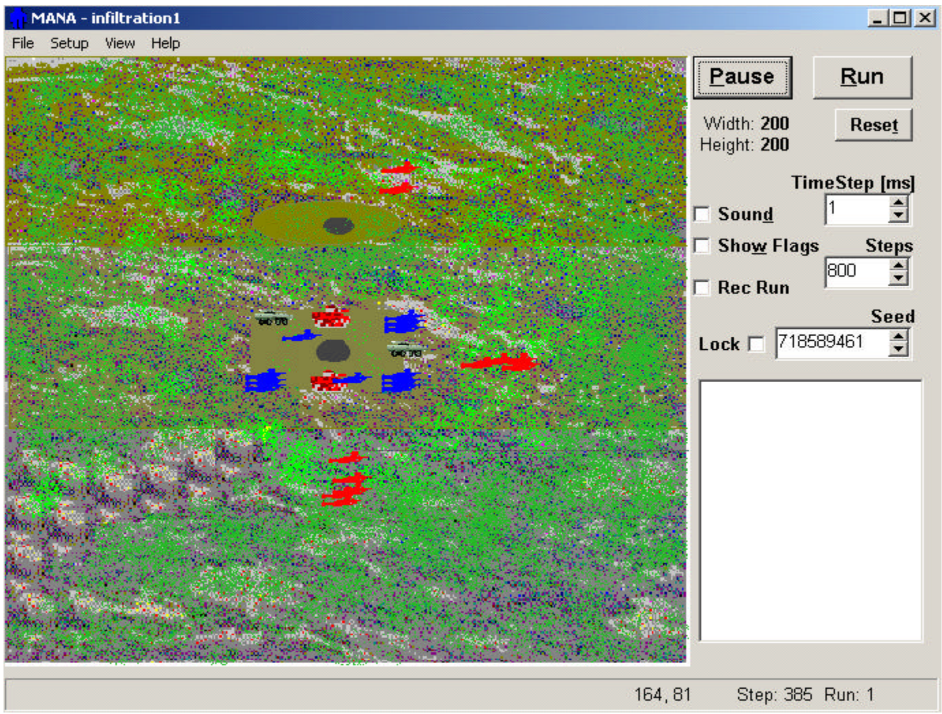

Figure 1. Infiltration Scenario (base-line) developed for use with MANA

The primary goal of this thesis is to explore this scenario in MANA and to create statistical models predicting Blue and Red casualties. To simultaneously investigate the effects of as many as 22 input variables, recently developed Near Orthogonal Latin Hypercube Designs and Fractional Factorial Designs are used. Utilizing a personal computer and the computational capabilities of supercomputers run by Mitre for the Marine Corps Combat Development Center (MCCDC), approximately 200,000 MANA runs were completed. The data are analyzed and graphically displayed using S-Plus, Clementine, Ggobi, and Netica Software packages. Several statistical models are developed and compared using a variety of diverse statistical techniques, including

Cluster Analysis, Neural Networks, Regression Trees, Linear Regression, and Bayesian Networks.

This thesis also presents an examination of some low-intensity conflicts and guerrilla battles that were fought in the second half of the 20th century. In this examination, the focus is on the human dimensions of the battlefield, the tactics, and the techniques, which are particularly unique to guerrilla troops. Following this close look at some guerrilla battles, the fundamentals, nature, and the tactics of guerrilla warfare are emphasized.

Interesting findings of this study include:

- Results (MOEs) are mostly affected by factors associated with the Red force. In particular, the Red team's stealth is the most important factor in terms of both Blue and Red casualties.
- More cohesive guerrilla forces who do not stay with their injured and form big groups when attacking the enemy get better outcomes in an infiltration operation.
- Guerrillas obtain the best results in terms of enemy casualties by being aggressive, attacking enemy vehicles/commanders and repelling enemy infantry. However, this personality also causes them more casualties.
- In this specific infiltration scenario, the Red side can negate the Blue side's advantage in firepower and numbers by using 16 to 27 guerrillas in each infiltration team.
- Regression Tree models, enhanced with MART, and Bayesian Network models are the most valuable analysis tools in this study. Specifically, they have the best visual representations of the data that are easy to interpret. Moreover, these models provide at least as good predictive power as all other models.

I. INTRODUCTION

"It is with the muscles of the intellect, with something like cerebral reflexes that the man of war decides, and it is with his qualities of character that he maintains the decision taken. He who remains in abstractions falls into formula; he concretes his brain; he is beaten in advance."

General Cordonnier (French Commander)

A. OVERVIEW

Since the end of WWII, a host of groups and states have pursued their interests in the Low Intensity Conflict environment. Many international wars and insurgencies have taken a heavy toll of lives and treasure [Ref.11]. Most of these wars have occurred in the Third World, changing the international environment. Many Third World conflicts were originated in the struggle to end the system of European empires. As nations achieved this goal, clashes among more or less conventional military forces sought to resolve artificially imposed relationships among newly independent states. This type of conflict continues. More frequently, insurgents have sought to alter the political, social and economic organization of these states, bringing about internal conflicts, which also continue. However, the means by which groups and nations conduct these conflicts have changed significantly, increasing the risks in the Low Intensity Conflict (LIC) environment [Ref.11]. One of the characteristics of LIC is that it is executed mostly by the rules of asymmetric war or guerrilla warfare.

Guerrilla warfare is a set of military tactics utilized by a group within a state or a local population in order to oppose the government or foreign occupying forces [Ref.5]. In essence, guerrilla warfare determines the shape of the battlefield in the Low Intensity Conflict environment and it remains a potent weapon for which armed forces increasingly must be prepared.

Recently, many analysts are realizing that conventional Operations Research tools based on rigorous mathematical equations and a detailed physical description of combat cannot provide a realistic picture of the complex and dynamic situations in which many

military operations, and particularly guerrilla-type operations, are conducted. In such operations, either war-fighting or peacekeeping, the participants have to interact with hostile or potentially hostile forces, by responding to their actions. In the process, a new situation or environment is created, which in turn triggers off new responses from both sides. Furthermore, war arouses some of the strongest human emotions, making it difficult to anticipate the behaviors of individuals in a command and control chain [Ref.4].

In 1914, F.W. Lanchester established some differential equations, now known as Lanchester Equations (LEs), to model battles. There have been many extensions to and generalizations of the LEs over the years. These types of models are called equation-based and are usually deterministic models. A primary drawback of equation-based models (EBMs) is that they do not deal well with the dynamics of interactions between the combating sides nor their reactions to each other's actions. Another serious challenge to EBMs is that the world is fundamentally nonlinear, and, consequently, many problems defy the traditional scientific approach of analysis by decomposition. The non-linearity of war-fighting means that a small perturbation to some critical elements (initial conditions) can profoundly alter the outcomes, thus, making reliable prediction extremely difficult, if not impossible [Ref.4]. Finally, EBMs have great difficulty in modeling the so-called intangibles of warfare, such as human emotions, aggressiveness, fear, anger, team cohesion and so on.

With the advent of complexity theory and its application to warfare studies, some researchers are viewing warfare as a complex adaptive system (CAS), which adapts, evolves and co-evolves with its environment [Ref.6]. In this sense, a guerrilla force also behaves like a complex system, not only interacting with its environment but also dynamically (and often nonlinearly) interacting between components of the system, namely the interactions between different levels of commanders and the commanders with the enemy. Conventional EBM are not equipped to deal with such ever-changing conditions and non-linearity. At best, one can use EBM to explore a range of possibilities, with a limited number of responses or contingency plans built into the war-gaming. This, however, is far from being adequate for describing the behavior of a system as it adapts and evolves with the changing environment.

Agent-based models (ABMs) offer an opportunity to analyze the above-mentioned complex problems by concentrating on the behaviors of and interactions between the participating entities instead of the performance of specific weapons or sensors [Ref.4]. This is particularly important in guerrilla warfare where one side usually dominates with respect to firepower. In other words, we shift our attention from analyzing the performance of pieces of equipment to how different modes of operation may alter the outcome of a combat or peacekeeping operation. Attention is focused on how a guerrilla force utilizes information and acts upon it, and the consequence on the success of a mission. In summary, the focus is on the emergent patterns of the *whole* rather than the individual *part*s.

The use of intelligent agents to study the emergent behaviors of complex systems and operations, such as the roles played by training, aggressiveness or the effectiveness of command-and-control in combat performance, has attracted much attention in the military operations research (OR) community in recent years [Ref.4]. Notable examples are models like the Irreducible Semi-Autonomous Adaptive Combat (ISAAC) [Ref.6] initiated by the US Marine Corps Combat Development Command (MCCDC) and the Map Aware Non-uniform Automata (MANA) [Ref.1] developed by the Defense Technology Agency (DTA), New Zealand.

B. PURPOSE AND RATIONALE

The purpose of this thesis is to utilize MANA as a scenario exploration tool with which to explore and exploit the non-linearity and intangibles arising from various interactions between Conventional and Guerrilla forces and their surroundings. Additionally, this study intends to find unsuspected links between a variety of factors in order to design a model that can explain these relationships reasonably well. We concentrate on the factors that affect the outcomes the most. These factors can be seen in the next chapter as well as Appendix E and Appendix F. We answer the following questions based on the results at the end of the analysis:

1. Which configuration is tactically more efficient for the Blue side, in the scenario used in the analysis, positioning vehicles up to the North, and infantries down to the South or locating the vehicles on the main directions, and infantries in between them?

2. In infiltration operations, guerrilla forces use single or multiple lanes to approach and attack the enemy's base. Which one is tactically more reasonable, in our scenario (see chapter III for scenario description), based on the success of the mission?

3. In our scenario, the Blue side has a significant superiority over the red in terms of the number of agents. How many agents are needed on the red side to overcome this deficiency?

4. In guerrilla warfare, the "men over weapon" concept steps forward in most situations. What significance does the Red side's personality changes (e.g., aggressiveness, cohesion, trust, fear, etc.), when engaging with the enemy, have on the outcome of the battle?

5. In case of a skirmish, the key objective for both sides, which leads to a successful operation, is the effective use of weapons. What is the optimum number of targets to be assigned to each member of the Red reconnaissance team when taking into account their firepower and fire range?

6. Under fire suppression many infantrymen tend to hide and take shots at the enemy without aiming at them. This kind of attitude leads to increasing stealth and decreasing sensor range. What consequences result, for both sides, under different circumstances when changing these parameters?

7. Unit cohesion can be perceived as the bonding together of members of a guerrilla force in such a way as to sustain their will and commitment to each other, their unit and the mission. Does unit cohesion have a significant affect on the outcome of the battle in a guerrilla-style war?

It is very important to realize that MANA is not intended to describe every aspect of a military operation. MANA's limitations must also be noted. Rather than having heavily pre-potted behavior representing a "perfect way" of doing things, we explore many different ways of accomplishing tasks quite quickly (due to the simplicity of the

model), even if they are "imperfect". After all, to be able to ask which kinds of formations appear to produce the best results, without placing constraints upon ourselves, is highly useful [Ref.1].

Following this introductory chapter, Chapter II provides the background motivation. Chapter III (supplemented with Appendix B) gives a detailed description of the scenarios, research questions, the model (MANA), and its parameters. Chapter IV explains the analysis methodologies used to explore the MANA data while chapter V explains the results of the analysis and uses these results to examine the research questions. Chapter VI provides a list of recommendations for MCCDC, for DTA, and for the analysts in improving the analysis of LIC and the development of MANA.

C. THESIS SCOPE

MANA has approximately 37 parameters, 30 of which can take different values for different trigger states. There are 10 trigger states. Roughly speaking, there are 307 parameters. To explore the effects of varying 307 parameters is a demoralizing task. For example, a two-level full factorial design requires $2^{307} = 2.6 \times 10^{92}$ runs to obtain one data point for each of the possible combinations. Even if a fractional factorial design is used it will not help us in this case. Therefore, 22 parameters are determined, which are believed *a priori* most likely influence the results. We also decided to have 129 levels of each parameter. With 22 factors and 129 levels we are not able to use factorial designs. Therefore, modified Latin Hypercube designs are used, which allow improved coverage of the input space. Thousands of preliminary runs were completed to explore many of the parameters in MANA. Following these preliminary runs, the model, together with the scenarios and the parameters that will be changed, were sent to Marine Corps Combat Development Command (MCCDC) for 153,900 runs. After further analysis of the results, the most important six parameters are chosen. Since the number of variables is within the computational capabilities of fractional factorial designs, this time a 3^{6-1} fractional factorial design is used. For the second time, the model, the base-line scenario and the design were sent to MCCDC for 24,300 runs. MCCDC formed a cluster of workstations, called "Gilgamesh", which provides analysts the ability to run some agent-

based models, like MANA, more than 100,000 times within a couple of days and to produce over 100,000 data points within 24 hours.

II. BACKGROUND

"There is no studying on the battlefield. It is then simply a case of doing what is possible, to make use of what one knows and, in order to make a little possible, one must know much."

Marshal Foch

A. CHAPTER OVERVIEW

Chapter II presents an examination of some low-intensity conflicts and guerrilla battles that were fought in the second half of the 20th century. In this examination, the focus is on the human dimensions of the battlefield, the tactics, and the techniques, which are particularly unique to guerrilla troops. Following this close look at some guerrilla battles, the fundamentals, the nature, and the tactics of guerrilla warfare are emphasized.

B. DEFINITIONS

This section gives definitions for some of the expressions used in this chapter.

Guerrilla literally means little war and derives from the activities during the Spanish national struggle against French occupying forces during the Peninsula War between 1808 and 1814 [Ref.5].

Guerrilla warfare or *asymmetric war* is a set of military tactics utilized by a minority group within a state or a local population in order to oppose the government or foreign occupying forces [Ref.5].

Revolutionary guerilla warfare or *insurgency* is a campaign fought by a minority group within a state to gain political power through a combination of rebellion, propaganda, and military action [Ref.5].

Counter-guerrilla warfare and *counter-insurgency* are the actions of existing governments and their armed forces to combat guerrilla warfare, insurgency and to prevent its resurgence [Ref.5].

Insurgency is an effective means either of achieving power and influence in a state or of bringing a particular cause to the notice of the national or international community [Ref.12].

Low intensity conflict is a political-military confrontation between contending states or groups below conventional war and above the routine, peaceful competition among states. It frequently involves protracted struggles of competing principles and ideologies [Ref.11].

C. HISTORY

Nuclear parity, the successes of deterrence, and an increasingly interdependent world have created a period of transition in diplomatic relationships, especially between superpowers, such as the United States of America (USA) and Russia. Regional powers have developed during this process, diffusing the international balance of power. Although the absolute strength of the superpowers has not declined, their relative strength in the world is less than it was two decades ago. Lesser powers have proliferated and have their own interests to pursue. Their independent actions provide many new possibilities for conflict, irrespective of relations among the superpowers [Ref.11].

In some respects, the end of the Cold War with the disintegration of the communist system in the Soviet Union between 1989 and 1991 was expected to decrease the likelihood of conflict in the modern world. Undoubtedly, the end of this war diminished the prospect of major global conflict and removed the external stimulus to many so-called "proxy" conflicts waged by the superpowers through groups in third countries. The Gulf War in 1991 suggested that in a "new world order" the international community, acting in concert, would promptly restore peace wherever it was threatened. The reality is that the conflict has not ceased. Between 1980 and 1990 there were approximately 60 conflicts inflicting 5.3 million deaths [Ref.37]. Whereas between 1990 and 1996 at least 98 conflicts occurred inflicting 5.5 million deaths—with only 7 of these conflicts being waged between recognized states [Ref.5]. Taken together, these factors reveal a world with a high potential for violent conflict. In other words, internal conflicts

and challenges to states from sub-national groups have not ended, and guerrilla warfare, or asymmetric war, remains as a potentially effective method to resolve these conflicts.

In this study, we look at six guerrilla battles from the 20th century with the intent of drawing some conclusions from them in terms of guerrilla tactics and techniques. The reason for choosing these six battles is that some of the important asymmetric war rules emerged from them. We intend to provide some insights to these battles, which are not known broadly. Brief summaries and lessons learned from these six asymmetric wars can be found in Appendix A. The following section gives a summary of the lessons learned from these conflicts.

D. COMMON LESSONS LEARNED FROM ALL SIX BATTLES

After analyzing these six battles it has been noticed that some of the lessons are common for all of the battles. These general conclusions are:

- "Winning hearts and minds" appears as the primary approach, leading to the success of the guerrilla operation.
- A small force can defeat a larger force if it achieves surprise and attacks the enemy where it least expects an attack.
- Most guerrilla operations are offensive, not defensive—seldom is there an attempt to seize and defend objectives.
- At the tactical level, one of the most important principles to be inferred from guerrilla operations is that guerrillas must be masters of their environment.
- These conflicts also emphasized the significance of night combat and combat under poor weather conditions and the importance of safe havens.
- These conflicts emphasized once again the need for a limit to the physical and mental strain that can be placed on guerrilla fighters.
- In all conflicts, the development of timely, accurate intelligence led to the success of tactical operations. This intelligence originated in large measure from the local inhabitants and police organizations. Good intelligence

appears to be a crucial cornerstone for guerrilla operations in counterinsurgency and low-intensity conflicts.

- For most of the conflicts analyzed so far, guerrilla forces seem to have time on their side. They prove to be more adaptable to tough conditions as opposed to anti-guerrilla forces. Therefore, guerrilla forces can stand all kinds of poor weather conditions, in rugged terrain, surviving for a longer time. This provides a big advantage in the long run. Infantry units can overcome this disadvantage by using guerrilla rules. In other words, whenever they can train themselves as guerrillas they win the battle in a short amount of time.

E. GUERRILLA TACTICS AND TECHNIQUES

It has become a cliché that "today's guerrilla is tomorrow's statesman, although someone's freedom fighter can just as equally be another's terrorist." Guerrilla warfare is one characteristic of an insurgency where the guerrilla is the combat element in the insurgency. When guerrilla forces first become operational, they usually engage in limited or small-scale operations. If they reach more sophisticated levels of organization, equipment, and training, larger operations using more conventional tactics may be expected. Insurgent tactics are characterized by elusiveness, surprise, and brief, violent action. Inevitably, guerrilla groups operate in difficult terrain, such as mountains, deserts, and forests. Within these environments, guerrillas possess local knowledge that is often denied their opponents. Often, they enjoy a degree of popular support from local inhabitants. Additionally, guerrillas are generally more mobile than their opponents and more prone to undertake "hit-and-run" raids enabling them to damage yet also evade their opponents, which thereby prolongs the struggle [Ref.11]. Guerrilla warfare can be also understood as a strategy of the weak faced by a stronger military power. By these means guerrilla warfare can be called as an asymmetric war. Guerrilla tactics in the early phases can be divided into two areas: terrorism and harassment.

1. Terrorism

The guerrilla can use terrorism to accomplish his goals. Terrorist techniques include bombings, assassinations, kidnappings, threats, mutilation, murder, torture, and blackmail. Not all guerrillas use terrorism as a tool. If terrorism is used, it is usually for coercion or intimidation. Terrorism may also be used to discredit the government by provoking the government into overreactions alienating the populace or demonstrating its inability to protect them [Ref.11].

a. Coercion

Coercion persuades individuals to act favorably in given situations toward the guerrilla or insurgent movement, such as convincing a local mayor to revise policies on the guerrilla or gaining passive support while at the same time redirecting resources to the insurgent movement.

b. Intimidation

Intimidation modifies behavior. Generally-speaking, threats or fear of harm are used either toward the individual or his family and friends. Intimidation induces the populace to silence or non-cooperation with government forces. It discourages competent citizens from accepting vital low-level government positions, such as the killing of servicemen to encourage draft evasion.

2. Harassment

Harassment keeps government forces on the defensive. If successful, it causes government forces to react to guerrilla operations. As a result, the government cannot conduct offensive operations that would prevent successful guerrilla operations. Harassment also weakens the government's resources and disrupts lines of communication. One advantage of harassment is the image it presents of the guerrilla being able to strike anywhere. Also, the government appears ineffective and incompetent by constantly losing or not clearly winning many small battles. This affects the morale of the government force and the populace [Ref.11].

The guerrillas use infiltration during movements. However, near the target area, small guerrilla elements mass and then conduct operations. The most common techniques employed by the guerrilla are the ambush, raid, and small-scale attacks. These are usually targeted against security posts, small forces, facilities, and Lines of Communication (LOC).

While government forces outnumber the guerrillas, the guerrillas seek to attain a local numerical superiority so that they can attain victory over small elements of the government forces. These tactics, if successful, compel government forces to commit larger elements to defensive tasks. Once government forces move to the defensive, they lose the initiative and become reactive. This allows the guerrilla force time and space to develop so that they can engage larger government forces with more conventional tactics [Ref.11].

F. THE NATURE OF GUERRILLA FORCES

There are four primary characteristics that distinguish guerrilla forces from regular forces. These are self-reliance, psychological advantage, mastery of the environment, and versatility. The most important of these characteristics is an attitude of self-reliance. Guerrilla forces in a number of ways exhibit this attitude of self-reliance. For example, guerrilla fighters typically demonstrate a strong will for surviving or succeeding in whatever situation they are found. They are fearless about unfavorable conditions. Furthermore, they are accustomed to austerity. They have learned to do without comforts and benefits that other soldiers consider to be necessities. These forces are not tied to a logistic lifeline. Their attitude of self-reliance leads them to use any available resource to sustain themselves or to improve their combat capabilities. Moreover, guerrilla forces do not give up. Even when outcomes seem inevitable, they stay in the fight and attempt to turn situations to their advantage. Their self-reliance is typified by self-denial, fortitude, resolve, and resourcefulness.

This attitude of self-reliance gives guerrilla forces a psychological advantage over their enemies. Confident in their abilities, guerilla forces normally consider themselves to be tactically superior to their opponents. Once they have demonstrated this tactical

superiority, their enemies often become fearful and wary. Guerrilla forces use this psychological advantage to keep their enemies off-balance and tense. Unpredictable, invisible to view, skilled with unanticipated methods, guerrilla forces can often paralyze the minds and wills of their enemies before a battle begins.

This self-reliant attitude enables guerrilla forces to be the masters of their environment. They do not fight, fear, or resist the environment; they embrace it as a shelter, protection, provider, and home. A guerrilla learns to be comfortable and secure in any terrain and climate, be it jungle, mountain, desert, swamp, or arctic territory. Exceptionally adaptable, guerrilla units dominate the terrain in which they operate and use it to their advantage against their enemies.

Because guerrillas understand and accept terrain and climate as their natural environment, they possess an unmatched tactical mobility on difficult ground. Moving with astounding speed and ease, guerrilla forces routinely use routes and traverse areas considered impassable by regular troops.

Mastery of the environment and the attitude of self-reliance give guerrilla units unusual versatility. Guerrilla units adapt quickly from one environment to another or from one type of operation to another. Holding a hilltop as a base one-day, they may be ordered to conduct a deep raid, mount a long-term reconnaissance patrol, participate in an ambush, or attack a fortified position on the next day. Their versatility is also reflected in a propensity for improvisation and innovation. Guerrilla forces naturally derive new tactics, if necessary, because they are not tied dogmatically to a specific doctrine. They use their equipment in innovative ways, without hesitating to use their enemy's weapons and resources when possible. Guerrillas also remain open to new ideas, new technology, and new weaponry. Guerrilla forces maintain a flexible attitude toward the battlefield.

G. THE TACTICS OF GUERRILLA FORCES

Guerrilla forces tend to hide and rest during the day and to move and fight at night. The vulnerability of guerrilla forces to enemy artillery and airpower forces them to use the cover of darkness for protection.

Guerrilla forces are best suited for offensive operations. Undoubtedly, the very character of the guerrilla forces is to be offensively oriented, and to retain the initiative in combat. Even when employed in an overall defensive strategy, a guerrilla unit constantly seeks opportunities to conduct offensive operations.

Guerrilla operations are usually conducted at very close range. Guerrilla forces normally do not seek to maximize the range of their weapons. Instead, they seek to get close enough to the enemy to smell and hear them.

The conventional tactics practiced by regular forces, significant combat support, massing of combat power, and large-unit maneuver do not work well for guerrilla forces. Instead, four main features characterize guerrilla tactics: surprise, shock, speed, and intelligence.

Guerrilla units achieve surprise in both time and space through several means. Through superior field craft and domination of the terrain, guerrilla units approach enemy positions with animal-like stealth. Moving at night, using every fold in the ground, exploiting every bit of concealment, and making no noise, guerrilla units frequently reach hand-grenade range of the enemy positions before they are detected. Guerrilla units also attack from unexpected directions and from more than one direction when feasible. Through pre-attack reconnaissance, guerrilla leaders determine the weaknesses and gaps in enemy dispositions, which then become the objects of attacks. Finally guerrilla forces vary the time and style of their operations. Thus, the enemy is unable to predict their actions.

Having achieved surprise, guerrilla forces shock the enemy with the speed and power of their attack. Although lightly armed, guerrilla units can deliver a heavy volume of fire for short periods of time by massing all its weapons forward in coordinated, accurate fire. The application of such heavy firepower, combined with rapid maneuver to the sides and rear of the enemy's positions, creates a violent shock effect possibly leading to a quick victory.

Guerrilla units also exploit speed in their operations. To achieve speed, which is a function of superior individual and group tactical movement, a guerrilla unit relies on its intimate knowledge of the terrain, a high level of fitness, expert field craft, and the

capability to negotiate difficult ground. By moving to an objective faster than its enemy thinks possible, guerrilla units can achieve surprise. To execute surprise successfully, however, requires stealth. While many units can only move rapidly if they make no attempt to conceal their movements, guerrilla units must accomplish stealth as well as speed if they want to be effective. Therefore, guerrilla units keep to the tough terrain areas and seldom use roads or trails.

Accurate, timely intelligence is vital to the success of guerrilla operations. To be effective, guerrilla forces must know what the enemy is about, while keeping the enemy in the dark about their own intentions.

Guerrilla units obtain tactical intelligence from a variety of sources. Often times, the majority of their intelligence comes through comprehensive patrolling. Guerrilla forces also utilize existing intelligence networks, rather than attempting to duplicate them. Therefore, smooth coordination with civil and police intelligence is absolutely essential in low-intensity conflicts, particularly counter-insurgencies.

The guerrilla forces also employ other local sources of information: guerrilla leaders use local guides, when necessary, in unfamiliar terrain. Sensitivity to intelligence remains an imperative for guerrilla operations.

In this chapter, an examination on six guerrilla battles from the 20th century, the fundamentals of guerrilla warfare, the tactics and the nature of guerrilla forces are provided with the intent to give the reader a better understanding of the essence of asymmetric war and the rules associated with it. In the next chapter, an infiltration scenario taken from personal experiences, the model and the parameters will be explained.

THIS PAGE INTENTIONALLY LEFT BLANK

III. SCENARIO AND MODEL DESCRIPTION

"It's the unconquerable soul of man, and not the nature of the weapon he uses, that ensures victory."

US World War II General, George S. Patton

A. CHAPTER OVERVIEW

This chapter starts with the description of an infiltration scenario, which is developed based on personal experiences in Turkey. There is a brief description of two additional scenarios, which have been developed for comparison purposes. This chapter also discusses the model, MANA in depth and explains the design philosophy behind it. It additionally addresses why the model is needed; why it can be superior in certain aspects to detailed models, like JANUS or CAEN; what improvements make it different from other Agent Based Models, like ISAAC; and why the chosen parameters were explored in the analysis. Appendix B provides additional details on MANA.

B. INFILTRATION SCENARIOS

1. Base-line Scenario

This scenario is based on personal experiences in Turkey. In September 1999, I was a platoon commander of a battalion stationed in one of the rural areas in the Southeast of Turkey. We received a mission order from the battalion commander to form a tank-infantry platoon and to move immediately to the top of a nearby hill. The platoon's task was to take up a position on the hill and protect the area against any terrorist activities. This location was very critical for both sides, because terrorists were using this territory as a passage from northern Iraq to Turkey. From the peak, we could easily monitor and interdict the terrorists. Consequently, we expected the terrorists to attempt sneaking into our base to damage assets, cause some casualties, and demolish our morale.

In the scenario, the Blue forces, which represent the friendly side, are composed of two tanks, two ACVs (Armored Combat Vehicle), and nine infantrymen. The platoon commander positions vehicles and personnel to watch the area and to protect assets against terrorist activities. To achieve this, the commander locates one tank up facing the North, the other tank down facing the South and one ACV right facing the East, and the other ACV left facing west while situating three infantry teams, each of which has three infantrymen, between the vehicles. The commander takes control of the northern portion of the base while ordering the non-commissioned officer (NCO) to take charge of the southern area. The Red forces, which represent terrorists, are composed of 11 guerrillas who are equipped with light infantry weapons. Two of these terrorists form a reconnaissance team. The remaining nine terrorists split into two infiltration teams, one from the South and the other from the East. The Blue force has superiority over the Red force in terms of weapons, equipment, and firepower. On the other hand, the Red side has the superiority in adapting to tough weather/terrain conditions, and in using surprise effectively.

In this specific scenario, the Red side conducts reconnaissance patrols against the Blue side for a long period of time until the Red commander makes sure that the necessary information and intelligence has been gathered about the Blue side to initiate the operation. This information gathering time can be extended from one week to one or two months. The Red side does not want to take any risk when unsure about the success of the attack. Therefore, the Red commander assigns two subordinates to watch the Blue base and observe every single activity the Blue force makes. These terrorists collect information about the Blue vehicles and infantrymen positions, the method of conducting their missions, the time of their shifts, and their strengths and weaknesses.

Once all of the needed information is collected, the Red commander commences the attack. The Red recon team initiates heavy fire upon Blue assets to distract its attention to that area. As soon as the Blue side reacts to the opponents' fire, the Red side initiates infiltration. The Red commander splits the forces into two groups and approaches the Blue base from two different directions. The commander prefers not to engage the enemy until reaching the target, being Blue's base and his critical assets--like command center, tanks, armored vehicles, etc. Time is another critical issue for the Red

side, which doesn't want to engage with the enemy for a long time, since engagement would probably cause many casualties and a failure of the mission. As a result, the objectives of the Red side are to establish sudden and heavy fire on the Blue side, to move quickly toward the enemy base, to destroy the critical enemy assets using grenades, bazookas, etc., to disengage with the enemy, and to flee from the area as quickly as possible. In contrast, the Blue side tries to suppress Red's initial fire keeping all attention on their field of responsibility. The Blue side tries to minimize friendly moves inside the base, to keep coordination and communication at the highest level, and to prevent the Red force from sneaking into the base.

After this general summary of the scenario we can conclude that the Blue side is greater in number, technologically more advanced, and well trained. Nevertheless, although the Red side is fewer in number and equipped with light weapons, their initiative, well-organized and informed approach, and quick reflexes keep them competitive. The screenshot below is the base-line scenario. This screenshot is best viewed in color.

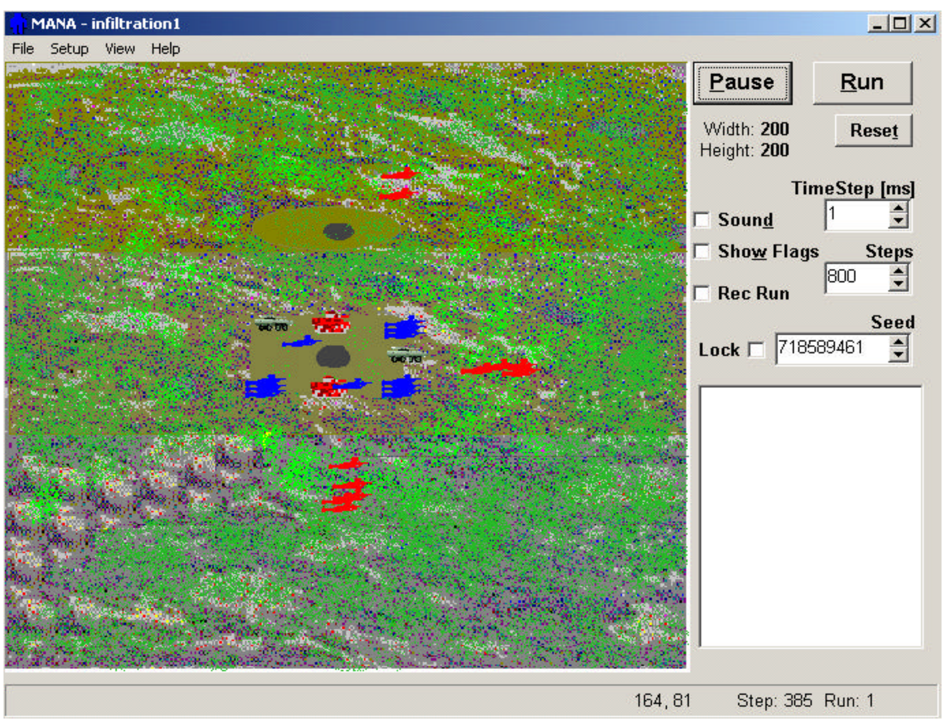

Figure 2. Infiltration Scenario (base-line) developed for use with MANA

2. Scenario Two

Scenario one was run with a different placement of the Blue agents. In this scenario, Scenario two, the Blue vehicles move up to the North while Blue infantrymen move to the South. There is no difference between the two scenarios in terms of any other parameter values. All of the agents' personalities and factor settings are assumed to be exactly the same in both scenarios.

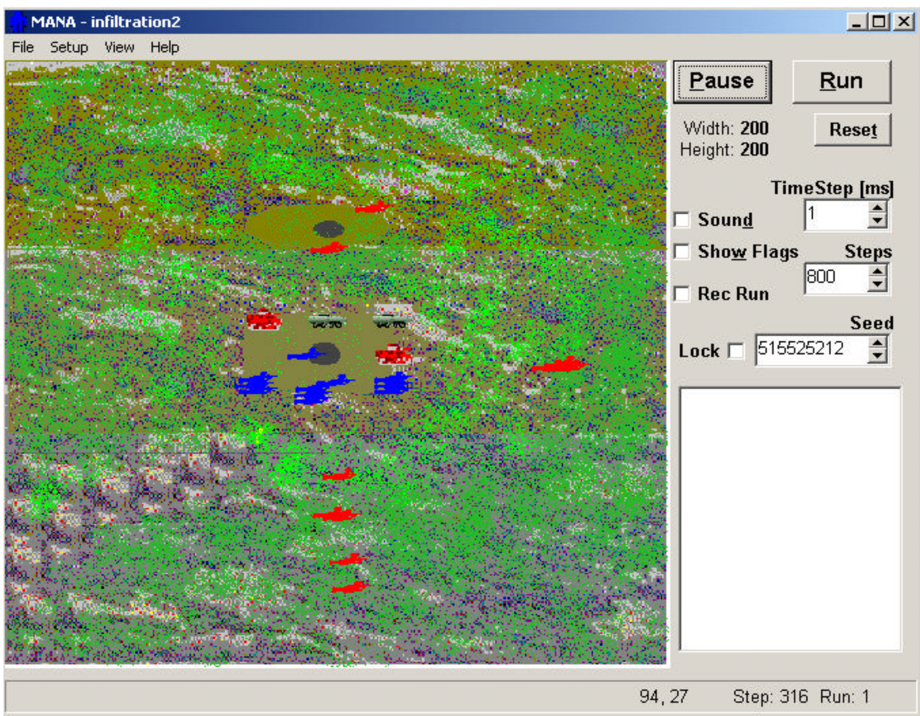

Figure 3. Alternative Infiltration Scenario developed for scenario comparison.

3. Scenario Three

Scenario one was again modified, this time with three infiltration lanes for the Red side, instead of two. In this scenario, the Red side uses another infiltration lane to sneak into the Blue base. In addition to the teams coming from the South and East, a third team approaches the enemy base from the west. Except for this third lane, there is no other change in any of the original parameter values.

One of our goals is to compare the second and third scenarios with the first one to identify any significant difference between the scenarios and to observe which one has

fewer casualties. This allows us to assess the effect of the Blue force's disposition and the Red force's tactics.

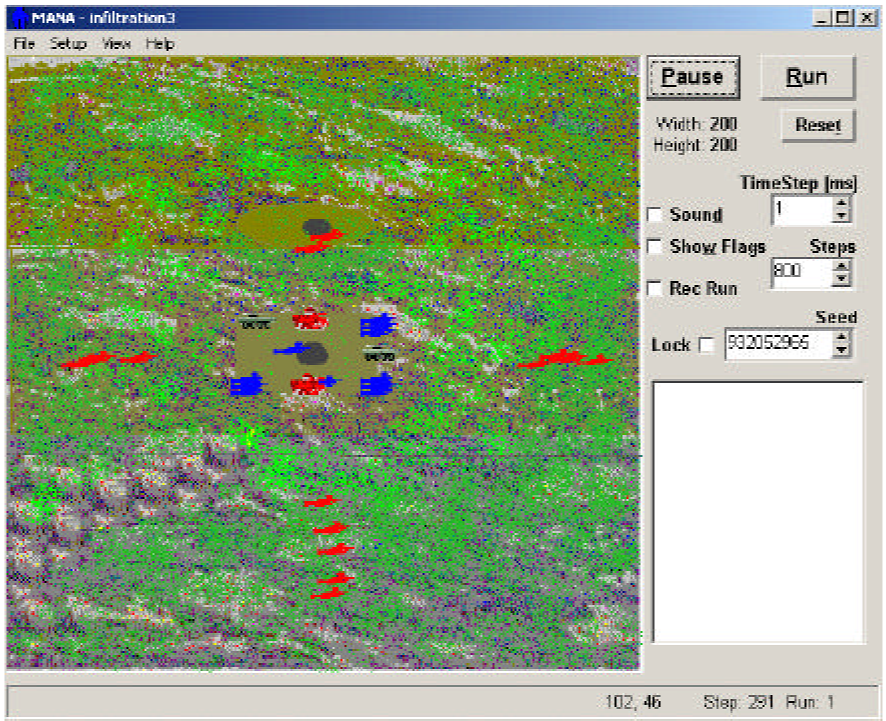

Figure 4. Alternative Infiltration Scenario developed for scenario comparison.

C. RESEARCH QUESTIONS

Following this experience of close combat, some questions emerged on the possible consequences of this skirmish. Agent Based Models, such as MANA, provide a great opportunity to answer those questions. The following are the questions that will be examined with MANA:

1. Which configuration is tactically more efficient for the Blue side, in this specific scenario: positioning vehicles up to the North, and infantries down to the South, or locating the vehicles on the main directions, and infantries in between them?
2. In infiltration operations, guerrilla forces use single or multiple lanes to approach and attack the enemy's base. Which one is tactically more reasonable, in this scenario, based on the success of the mission?

3. In this scenario, the Blue side obviously has a significant superiority over the Red side in terms of the number and the ability of agents. How many agents are needed for the Red side to overcome this deficiency?
4. In guerrilla warfare, the “men over weapon” concept steps forward in most of the situations. What significance does the Red side’s personality changes (e.g., aggressiveness, cohesion, trust, fear, etc.), when engaging with the enemy, have on the outcome of the battle?
5. In case of a skirmish, the key objective for both sides, which leads to the success of the operation, is the effective use of weapons. What is the optimum number of targets to be assigned to each member of the Red reconnaissance team considering their firepower and fire range?
6. Under fire suppression many infantrymen tend to hide and take shots at the enemy without aiming. This kind of attitude leads to an increase in stealth and a decrease in sensor range. What are the consequences of changing these parameters?
7. Unit cohesion can be perceived as the bonding of members of a guerrilla force in such a way as to sustain their will and commitment to each other, their unit and the mission. Does unit cohesion have a significant affect on the outcome of the battle in a guerrilla-style war?

D. OVERVIEW OF MANA

The following sections taken primarily from MANA User’s Manual [Ref.1] describe the model and its features. Map Aware Non-uniform Automata (MANA) was developed by New Zealand’s Defense Technology Agency (DTA).

- **Map Aware** - Agents are aware of and respond to, not only their local surroundings and terrain, but also a collective registry of recorded battlefield activities.
- **Non-uniform** – Not all agents move and behave in the same way.
- **Automata** – Agents can react independently to events, using their own personalities.

E. THE NEED FOR MANA

The history of physics has been characterized by the search for systems simple enough to be described with a high degree of accuracy by mathematical equations [Ref.1]. Isaac Newton's laws of motion are an example. Although extremely accurate at predicting, for instance, the path and distance traveled by a heavy projectile, the laws cannot in general be relied on in all circumstances. If the projectile is light, it becomes subject to a far greater degree of viscous drag in the atmosphere, which makes the original calculations invalid. These equations cannot, with complete generality, be simply or easily corrected for the action of the drag. The reason is that the interaction between the viscous atmosphere and the projectile is just too complicated while depending on too many variables (particularly if the projectile is light and has irregular shape, like a feather), not to mention that the atmosphere itself is unpredictable and turbulent [Ref.1].

This simple example illustrates a powerful point: the world is far more complicated than Newton's equations. To this day, no set of equations exist that can with absolute certainty predict the evolution of many of phenomena seen in everyday life for any significant period into the future [Ref.1].

Therefore, to rely on models built "on a bedrock of physics" is deceptive. It is a myth that a more detailed model is necessarily a better model, because it is impossible to capture accurately every aspect of nature. In fact, the more detailed a model is, the more obscure its workings. This problem is compounded if the user is not the model designer [Ref.1].

The non-linear nature of equations describing many real world phenomena makes them extremely sensitive to initial conditions. This means that even infinitesimal errors in describing the real world initial conditions may cause the model to make predictions that are almost uncorrelated with actual events [Ref.1]. Many circumstances occur in which events unfold with such sensitivity to initial conditions that predicting the outcome of a single event is impossible. Although little can be done about this non-linearity, knowing when it arises can be critical [Ref.2]. Awareness of non-linearity helps identifying and avoiding possible problems. Situations also arise in which a very small

change can dramatically alter the likely outcome. In this case, efforts should be concentrated on the source and condition in which non-linearity may arise.

Another concern of most military analysts is the intangibles associated with war. Predicting battlefield performance, an element of war, is most difficult to measure or count because it depends so heavily on human behavior [Ref.2]. Cohesion, discipline, leadership, motivation, trust, and fear are accepted as important factors of intangibles.

Moreover, in the case of the important elements of guerrilla warfare, conventional equation-based or low-resolution models are incapable of directly modeling many important aspects of guerrilla warfare. Studying the effects of non-linearity and intangibles, inherent in guerrilla warfare, on the outcome of the battle is where Agent-based models (ABM) may be more effective than equation-based or low-resolution models.

F. ENTER MANA

MANA was designed for use as a scenario-exploring model. It is intended to address a broad range of problems [Ref.1].

MANA is based on two key ideas:

- The behavior of the entities within a combat model (both friend and foe) is a critical component of the analysis of the possible outcomes.
- Time is wasted on highly detailed models for determining force mix and combat effectiveness.

The behavior of troops in any kind of scenario plays an important role. However it is often overlooked by analysts because human nature is a mathematical intangible; just as is the weather. As with many other intangibles in analytical combat models, these behaviors are often thrown away in favor of dwelling on infinitesimal details, often of little real importance to a given scenario [Ref.1].

It is precisely the aspects that are most difficult to describe which are of the greatest importance. So much in war is intangible, and these intangibles are spoken about

so often it is impossible to ignore them [Ref.1]. (See Muzaffer Coban's thesis [Ref.41] for a good example of this argument).

G. REASONS FOR USING MANA

MANA is not intended to describe every aspect of a military operation. Furthermore, there is no inbuilt "intelligence" which determines the plan the MANA entities are working towards [Ref.1]. When setting up any scenario in the model, one should be very cautious. There must be a clear idea of which aspect of warfare the scenario is addressing, and what the entities are trying to do. The non-linear nature of the model ensures that, regardless of the modeler's presumption, a remarkably large number of outcomes are possible. Such a range of outcomes is characteristic of complex adaptive systems, and occurs even with quite simple rules of behavior [Ref.1].

The limitations of MANA must also be noted. The model is not designed to carefully examine formation fighting, for example. Formations occur in MANA as much by accident as by design. Furthermore the entities within MANA do not always behave in a sensible manner. They often exhibit actions that might be described as "mistakes". However, this is part of the point of MANA [Ref.1]. Exploring the different ways of doing things quite quickly is an option (due to the simplicity of the model), even if they are "imperfect". This is highly useful when asking about the kinds of formations that produce the best results and what the consequences are in making certain kinds of mistakes, without placing constraints upon ourselves [Ref.1].

H. CELLULAR AUTOMATON MODELS

Descriptively, MANA is in a general class of models called Agent-Based Models (ABM). ABMs have the characteristic of containing entities controlled by decision-making algorithms. Hence an ABM contains entities representing military units that make their own decisions, as opposed to the modeler explicitly determining their behavior in advance [Ref.1].

The MANA model falls into a subset of cellular automaton (CA) models. CA models have their origin in physics and biology. These models are often called complex adaptive systems (CAS) because of the way the entities within them react with their surrounding [Ref.1]. Some properties of MANA and CAS combat models generally are

- They exhibit "global" behavior, which "emerges" as a result of many local interactions.
- They are an example of a process of feedback that is not present in "reductionist", top-down models.
- They cannot be analyzed by decomposition into simple independent parts.
- Agents interact with each other in non-linear ways, and "adapt" to their local environment.

The MANA model is an attempt to create a complex adaptive system for some important real-world factors of combat such as [Ref.1]

- Change of plans due to the evolving battle.
- The influence of situational awareness when deciding an action.
- The importance of sensors and how to efficiently use them.
- The effects of intangible personality factors, like fear and aggression.

I. DIFFERENCES FROM OTHER AGENT-BASED COMBAT MODELS

The MANA model builds on and complements the earlier ISAAC/EINSTEIN CA models developed by the Center for Naval Analyses, and the more recent but still incomplete, Archimedes model being designed for the US Marine Corps [Ref.1].

The primary use of MANA is as a "distillation" tool: creating a bottom-up abstraction of a scenario that captures just the essence of a situation, while avoiding unessential detail. Additionally, MANA was designed to explore key concepts that ISAAC (at that time) was unable to explore [Ref.1], in particular

- **Situational Awareness:** Includes a group memory of enemy contacts (also the method MANA uses to simulate communications).
- **Terrain Map:** Contains roads that agents can follow.
- **Waypoints:** Defines a set of waypoints, not just an ultimate goal.
- **Event-driven personality changes:** Identifies events, such as being shot at, taking a shot, reaching a waypoint, making enemy contact that can trigger a different personality set lasting for a certain time. Personality changes include individuals or a whole squad.

J. USES OF THE MODEL

The MANA model explores the interaction of autonomous entities, referred to generally as automata or agents. Agents are the entities that appear in the model. For instance, in our scenario, Blue infantries, vehicles and Red guerrillas are referred to as agents in MANA. Each entity has certain personality traits that drive it toward or away from other entities on the "battlefield". In MANA, agents have the ability to change their personality completely if certain events occur [Ref.1].

Parameters within MANA can be divided into three basic types. The first, personality weightings, determine an automaton's propensity to move towards friendly or enemy units, its waypoint, an easy terrain, and a final goal point.

The second type, constraints, acts as conditional modifiers to this process. The Cluster parameter "turns off" the automaton's propensity to move towards friends above some maximum cluster size; the Advance parameter prevents an automaton from moving towards its objective without a minimum number of friendly units accompanying it. The Combat parameter determines the minimum local numerical advantage that a group of automata require before approaching the enemy [Ref.1]. A final set of parameters describes the basic capabilities of automata, such as weapon range, sensor range, movement rate, single-shot kill probability, defensive factor, stealth, and the maximum number of simultaneous targets that can be engaged. The automata also possess the

ability to communicate the position of enemy automata to other friendly automata. With these simple parameters, a surprisingly wide set of behaviors can be induced [Ref.1].

The meanings and usage of the parameters are explained in Appendix B together with the Graphical User Interfaces (GUI) associated with them. In the next section, only the parameters chosen for analysis purposes will be explained.

K. ACTIVATED PARAMETERS

In order to explore our research questions within the capabilities of MANA, we chose 22 parameters to vary for further analysis and development of our model. This section gives a general description of those parameters and the reasons they were chosen. In essence, we would like to give the reader a better idea of these parameters and their connections with our scenario. Prior to getting into the details of the parameters, we will review the reasons why we chose certain states for different agents. The formatting for the following parameters is Squad / State / Parameter. These are the states and the parameters chosen for different squads of the scenario:

- **Red Team / Enemy Contact:** We selected "enemy contact" state for Red infiltration teams because it represents the situations of terrorists the best. All terrorists reflect the same behaviors as soon as the enemy enters their sensor range. Moreover, this state of Red infiltration teams is the one that primarily affects the outcome.

- **Red Recon Team / Shot At:** We chose "shot at" state for Red recon agents mainly because their mission is to fire upon the enemy base for a certain amount of time to assist the infiltration. They are the ones who start the attack with their fires. This state best describes these agents' situation.

- **Blue Team / Taken Shot:** We selected "taken shot" state for Blue agents mainly because they are the defending side in this specific scenario. They are fired upon first and then they react to the fire.

- **Red Team / Number of agents:** This parameter is chosen to find out the best size for Red teams in order to compromise their deficiency in terms of

number. In the original scenario the Red force has 11 agents, but we want to investigate the size needed to turn the battle to the Red team's favor.

- **Red Team / Enemy Contact / w1:** This parameter controls agent's propensity to move toward agents of same allegiance in the state of enemy contact. We vary w1 parameter to see various effects of cohesion and proximity among agents.
- **Red Team / Enemy Contact / w2:** This parameter controls agent's propensity to move toward agents of enemy allegiance. We vary w2 parameter to see the effects of aggressiveness and fear of the agents.
- **Red Team / Enemy Contact / w3:** This parameter controls agent's propensity to move towards injured agents of same allegiance. Since terrorists never want to leave their injured or dead friends behind, we wanted to identify what the pay off would be for such a tendency.
- **Red Team / Enemy Contact / w8:** This parameter controls agent's propensity to move toward enemies in the SA map, which are of threat level 1. Since we set the threat levels of Blue infantrymen (excluding commanders) to 1, this parameter shows Red agent's tendency to move toward Blue infantries. In the real battle, the terrorists exhibited a tendency to move toward infantries. This parameter is also an indication of aggressiveness and fear.
- **Red Team / Enemy Contact / w10:** This parameter controls an agent's propensity to move toward enemies in SA map, which are of threat level 3. Since we set the threat levels of Blue vehicles to 3, this parameter shows a Red agent's tendency to move toward Blue vehicles. In the real battle, the terrorists demonstrated a tendency to move away from vehicles. This parameter is also intended to be an indication of aggressiveness and fear.
- **Red Team / Enemy Contact / Stealth:** This parameter shows the probability of the agent not being seen. In infiltration operations this

factor is one of the most vital elements of the operation. Infiltration teams always prefer to engage with the enemy at the closest distance available. Therefore, remaining invisible until they reach their target is important.

- **Red Team / Enemy Contact / Firepower:** This is simply single shot kill probability. Since terrorists have a disadvantage of firepower, efficiently using this firepower is very important.
- **Red Team / Enemy Contact / Combat Constraint:** The combat constraint permits advancing toward the enemy only if a certain numerical advantage is met, i.e., the Red side outnumbers the Blue side by a number of agents. This parameter also indicates how aggressive the Red agents are.
- **Red Team / Enemy Contact / Movement Range:** This is the number of grids an agent can move per time-step. In essence, this parameter indicates how quickly an agent can move in certain territory.
- **Red Team / Enemy Contact / Cluster Constraint:** The cluster constraint is designed to prevent agents from clustering in groups larger than a specified size, determined by the value of this parameter. Terrorists usually split into teams of two or three people. This constraint is important in discovering whether or not this formation is the right way of approaching the enemy.
- **Red Recon Team / Shot At / Stealth:** Critical to the terrorists' safety is their ability to remain hidden since they will be the targets of initial Blue fires.
- **Red Recon Team / Shot At / Firepower:** Since these terrorists initiate fire, their accuracy in shooting seems to be very critical for Red mission success.
- **Red Recon Team / Shot At / Sensor Range:** An agent can identify any other agent or terrain within this number of cells. Recon agents' positions and proximity to enemy base are very decisive in the success of their

mission. By improving this range the Recon team can collect more information about the Blue team and assist with the infiltration.

- **Red Recon Team / Shot At / Firing Range:** An agent can shoot at any other agent within this number of cells. The weapon selection for recon agents is also very important. They want to fire upon all enemy agents in the base so that their fires distract and harass all Blue agents.
- **Red Recon Team / Shot At / Max Targets per Step:** This determines the number of targets within both sensor and firing range that can be shot at in a single time step. Since the Red team wants to distract all Blue agents' attention towards the Recon team, this parameter is very critical. By changing this parameter, we can influence the performance of recon team, as well as the fate of the infiltration.
- **Blue Infantry / Taken Shot / Stealth:** As mentioned in one of our research questions, in guerrilla type skirmishes, infantries tend to hide and shoot at the enemy without aiming at him. By varying this parameter we are able to monitor this effect on the outcome.
- **Blue Infantry / Taken Shot / Firepower:** The attitude explained in the previous paragraph would obviously decrease the agent's probability of hitting and killing the enemy. We will play with this parameter to see its consequences on the outcome.
- **Blue Infantry / Taken Shot / Sensor Range:** The same effects are seen on the agent's sensor range when hiding and shooting aimlessly; therefore, our intention here by playing this parameter is the same with the previous two parameters.
- **Blue Infantry / Taken Shot / Firing Range:** Since by hiding these agents are decreasing their weapons' effective firing range, we want to observe this effect as well.
- **Blue Vehicles / Taken Shot / Sensor Range:** Although the Blue vehicles have a better sensor range, they were unable to use it mostly due to terrain

conditions: area is much wooded and the slope is very high. We hope to find certain patterns by playing with this parameter. The changes on this parameter can be a consequence of better positioning of the vehicles or reducing the effects of bad terrain conditions.

- **Blue Vehicles / Taken Shot / Firing Range:** Our intention by playing this parameter was the same as sensor range parameter of Blue vehicles.

L. TECHNICAL DETAILS OF THE MODEL

In MANA, each agent's artificial intelligence (AI) has been restricted as much as possible. The emphasis is on modeling the interactions of many simple agents rather than generating sophisticated behavior for individual agents [Ref.1].

MANA was developed in the object-oriented programming language, "Delphi". Agent based models are well suited to object-oriented development methods due to a tight correspondence of the entities to be modeled and the software objects. An object-oriented design also facilitates changing the model. The model consists of a few, key classes (of objects), which are described below [Ref.1]:

1. Battlefield Object

The Battlefield object contains an overview of the rectangular grid that represents the battle space where the agents interact. It has a map of the terrain and keeps track of the squads and all the agents. This object is responsible for moving all the agents, for adjudicating combat between agents, and for recording progress data for model output (measures of effectiveness). Additionally, it prevents agents from moving off the grid, into impassable terrain, or two agents occupying the same location. At each time step the battlefield object decides on an order for agents to move and fire and then calls each agent object to move itself [Ref.1].

2. Squad Object

The squad maintains a group of agents [Ref.1]. It also has a set of properties belonging to that squad that can be shared by the agents, such as

- A shared set of personalities that the squad agents can take on (and the events which trigger the personalities),
- A situational awareness map of enemy contacts, and
- Waypoints, which become the agent's goals, one after the other.

3. Agent Object

The Agent object knows where it is, and what state it is in, and how to detect the environment, fire and move. It has certain personality weightings to move toward or away from other objects. This object has the ability to modify its personality if certain local conditions are met, but not its entire personality set, which is the job of the parent squad [Ref.1].

4. Movement

The most important action of an agent is to move. The movement algorithm selects the grid square within its movement range that most satisfies its desire to move towards some entities and away from others. The current location is also an option, so the agent can stay put [Ref.1].

The movement algorithm consists of the following steps:

- Consider all moves within the movement range of the agent, including staying in place.
- Eliminate moves into locations containing other agents or impassable terrain.
- Consider all the entities in range by deciding on the most appealing of the permissible moves, using personality weights that represent a desire to move toward or away from agents, the waypoints, terrain or contacts on the Situational Awareness map.

- Impose behavior modifiers that change the basic behavior (e.g. minimum distance to others, cluster constraints, and so on). If a number of moves are nearly equal, then choose a move at random from the attractive moves.

Local sensor information takes precedence over information available to the agent from the situational awareness map. If an enemy is within sensor range, then the influence of the situational awareness is ignored [Ref.1].

a. The penalty calculation

The penalty calculation finds the move with the least "penalty". Moves are possible to grid squares within "movement speed" squares of the current location, which are not already occupied by an agent or impassable terrain [Ref.1].

If several moves have a similarly low penalty, a move is chosen at random from the good moves. The "movement precision" parameter sets how wide the margin should be for accepting similarly good moves. Setting the movement precision to a low value means that only the best move will likely be chosen and the movement will appear very deterministic. If the movement precision is too great, the agents tend to wander about in a Brownian motion, as moves are selected almost entirely at random [Ref.1].

The tendency to move toward or away from an entity is constant with distance. For example, the weighting to move towards the next waypoint is the same whether the entity is three cells or 150 cells away from it. The penalty for moving to any grid location is the sum of 10 penalty calculations, corresponding to the 10 personality parameters listed in Table 9 in Appendix B.

The algorithm used to calculate the penalty for a collection of entities within sensor range is the same for all 10 components [Ref.1]. The general algorithm for calculating the penalty component associated with a candidate move is shown in Figure 4. The important term in the algorithm is (NewDist + (100 - OldDist)) / 100.

This treats all entities as if they are about 100 units away. If NewDist < OldDist and a move closer is desirable, the penalty term will end up slightly less than 1.0.

If NewDist < OldDist and a move away from the entity is desirable, a penalty term of slightly greater than 1.0 will result, i.e. there is a greater penalty [Ref.1].

```
Given a candidate move...
Loop through all entities within range of agent's current location
      OldDist = units from current location to entity
      NewDist = units from candidate location to entity
      If NewDist < Minimum then Direction = -1 else Direction = 1
      Sum = Sum + Direction * (NewDist+(100-OldDist))/100
      NumEntities = NumEntities + 1
      Repeat loop
Penalty  = Sum / NumEntities
```

Figure 5. Penalty Calculation Algorithm

The penalty for moving towards or away from other agents is normalized by the number of agents (the last line of Figure 4). For example, if agents are attracted to friendly agents, they minimize their average distance to friendly agents within the sensor range. However, a number of constraints can modify this attraction. The minimum distance to friends/enemies range sets a radius to other agents, inside which the penalty is negated. The check against minimum distance is made for every agent within sensor range. If agents are within the minimum distance, their penalty is negated before adding to the sum for the friends' penalty component [Ref.1].

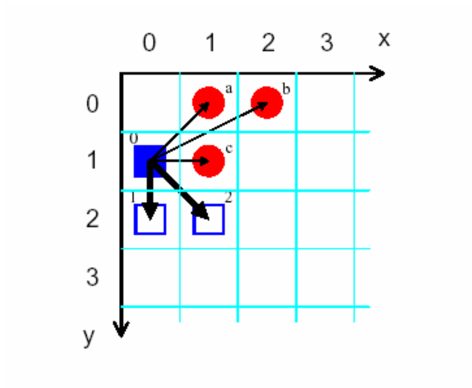

Figure 6. Penalty Calculation Diagram

5. Random Numbers

The random numbers generated in MANA are obtained using the built-in Delphi function "Random". Delphi uses a pseudo random number generator (PRNG) with a cycle of 2^{32}. It maintains a 32 bit seed, which is treated as an unsigned integer. The result can take on approximately 4.3 x 10^9 values (2^{32}) before repeating the cycle [Ref.1].

When the MANA application is started, the Delphi function, Randomize, is called setting the random number generator with a random seed obtained from the system clock. This can be overridden by setting the seed manually via the multi-run dialog or in the Seed box on the main window [Ref.1].

The Random function is used many times during the execution of the MANA model, for example in determining moving and firing order, in calculating stealth, in deciding the best of similarly good agent moves, in getting shot (SSKP), and in placing agents at the start of a run [Ref.1].

In the MANA source code, the Random function is normally called to return an integer number in a certain range. A typical example would be the firepower calculation. In the agent "get shot" function, a random number between 0 and 100 is calculated. If this number is less than the SSKP (an integer between 0 and 100) of the enemy agent who shot, then the receiving agent is hit. As one would expect, a high SSKP (close to 100) results in receiving agents being shot most of the time [Ref.1].

In this chapter the scenarios, the model and the parameters were described in detail. In the next chapter, analysis tools and techniques to explore the model and scenarios with specified parameter settings will be explained. Several statistical methods and packages will be used for the analysis. We will look at these tools and discuss the approaches used to explore and exploit the data.

IV. ANALYSIS METHODOLOGY

"To fight and conquer in all your battles is not supreme excellence; supreme excellence consists in breaking the enemy's resistance without fighting."

Sun-Tzu

A. CHAPTER OVERVIEW

This chapter explains the analysis methodology, designs and statistical tools used to explore small-unit guerrilla warfare in MANA. Our objectives are to develop descriptive and predictive models for the data, to display the data using visual techniques, and to find reasonable and plausible answers to our research questions. In this chapter, the measures of effectiveness (MOEs) and statistical designs used to evaluate the significance of the parameters in MANA are explained. This chapter also explains the statistical tools and the statistical techniques used in assessing the differences between scenarios, and the significance of the MANA parameters. In our analysis, we developed many different models using several statistical techniques. The reason for this is to compare these techniques and assess their descriptive, predictive and visual strengths and weaknesses.

B. MEASURES OF EFFECTIVENESS

The analytical features in the current version of MANA are designed only for Measures of Effectiveness (MOEs) based on the number of casualties at the end of a run. Therefore, in this study, two MOEs are used to examine the success of the mission for both sides and provide some insight into the seven basic research questions discussed in the previous chapter. The first MOE is the proportion of Red MANA agents killed while the second MOE is the proportion of Blue MANA agents killed during the battle.

1. Proportion of Red Agents Killed (MOE 1)

Undoubtedly, the most important objective of a guerrilla operation is maximizing the enemy losses while minimizing the friendly losses. Hence, the number of agents killed is always a critical factor when evaluating mission success.

Proportions of Red agents killed are used instead of actual values. The reason for this is that one of the 22 parameters chosen is the number of Red agents. Since the number of Red agents changes and takes different values in each run we realized that taking the actual values of Red agents killed may give misleading conclusions. Therefore, the number of Red agents killed is divided by the number of Red agents in each run with these proportions being used as one of our MOEs.

This MOE is used to explore all the questions at hand. First, it is used to compare the alternative scenarios to the base-line scenario. Sign Tests are completed based on the values of this variable. Then, the MOE is used to determine the statistical significance of the chosen 22 MANA parameters (for a complete list of these factors and their ranges see Appendix E). Although this MOE is considered in all 22 parameters, some parameters are more closely related to this MOE than others. The number of Red agents, Red team w1, Red team w2, Red team w3, Red team w8, Red team w10, Blue infantry firepower, Red team stealth, Recon team stealth, Blue infantry sensor range, Blue infantry firing range, Blue vehicles sensor range, Blue vehicles firing range, Red team combat constraint, Red team movement range, and Red team cluster constraint are the parameters that were specifically chosen to see their effects on MOE1. See Appendix E and Chapter III for more details.

2. Proportion of Blue Agents Killed (MOE 2)

In order to be consistent with MOE1, this MOE is chosen to be a proportion of the actual values as well. This MOE is also used to explore all of the research questions. It is used for comparison purposes between scenarios. A second set of Sign Tests is accomplished for the proportion of Blue agents killed. The significance of the 22 parameters is determined, for the second time, based on the values of Blue agents killed. As mentioned earlier for MOE1, MOE2 is also believed to have connections with all 22

parameters. However, these connections are stronger for some parameters; such as the number of Red agents, the Blue infantry stealth, the Recon stealth, the Recon firepower, the Red team stealth, the Red team firepower, the Recon sensor range, the Recon firing range, the Recon max targets per step, and the Red team combat constraint.

C. LATIN HYPERCUBE DESIGNS

When several factors are to be studied collectively, an experimental design is necessary to compare those variables. Experimental design is an efficient way to estimate the effects of several factors simultaneously. An experiment with at least one observation for every possible combination of levels is referred to as a complete layout (or full factorial design) [Ref.30]. If the number of factors is too many, like in our data set (22 factors), studying the effects of these factors with even two levels is a very challenging task. For instance, if we want to have a full factorial design with all the main effects and interaction terms then we need 2^{22} (4,194,304) runs to obtain one data point for each of the possible combinations. Even this number, without considering replications, takes a significant amount of computer time to be run. Assuming we want to use a half-fraction fractional factorial design instead of a full factorial design, we need 2^{22-1} (2,097,152) runs. After taking into account the replications (say the Central Limit Theorem thumb rule of at least 30), the resulting number is frequently either impractical because of cost, time, or space constraints, or literally impossible.

Furthermore, it is impossible to capture non-linearity with only two levels of each factor. A better space filling design with many levels of the factors is required to capture non-linearity well. For these reasons Latin Hypercube Designs were chosen.

Latin Hypercube sampling (LHS) was developed by W. J. Conover of Texas Tech University. His primary intention was to develop a method for improving the efficiency of simple Monte Carlo sampling. This sampling method can be thought of as stratified Monte Carlo sampling [Ref.32].

The method, LHS, uses a stratified sampling scheme to improve the coverage of the input space [Ref.33]. It selects n different values from each of k variables $X_1 \ldots X_k$ in the following manner. The range of each variable is divided into n non-overlapping

intervals on the basis of equal probability. One value from each interval is selected at random (or by a fixed method, such as the median) with respect to the probability density in the interval. The n values obtained for X_1 are paired in a random manner (equally likely combinations) with the n values of X_2. These n pairs are combined in a random manner with the n-values of X_3 to form n triplets and so on, until n k-tuplets are formed. This sample (or any random sample of size n) is easily imagined as forming a (n x k) matrix of input values where i^{th} row contains the specific values of each of the k input variables used on the i^{th} run of the computer model [Ref.33].

To fully appreciate the value of the underlying structure of LHS, it is helpful to be familiar with computer models used in actual applications [Ref.32]. Such models are usually characterized by a large number of input variables, and usually only a handful of these inputs are important for a given response. In addition, the model response is frequently multivariate and time dependent. If the input values were based on a factorial design, each level of each factor would be repeated many times. Moreover, the experimenter usually has a particular response in mind when constructing the factorial design; this design may be totally ineffective with multiple responses [Ref.32].

On the other hand, LHS ensures that the entire range of each input variable is completely covered regardless of which single variable or combination of variables might dominate the computer model response(s) [Ref.32]. By sampling the entire range, each variable has the opportunity to show up as important, if it indeed is important.

Also LHS is more efficient than simple random sampling in a large range of conditions [Ref.32]. The most basic method of collecting and monitoring data is simple random sampling. With this design, samples are selected randomly and with equal probability. While this method is easy to implement, a variety of sampling designs can be more efficient. One of those designs is LHS, which produces estimates with smaller standard errors for the same sampling effort, or requires fewer samples to obtain the same standard error possibly obtained with simple random sampling. Finally, the LHS method can cope with many input variables and is computationally cheap to generate.

A Latin Hypercube Design (LHD) has a few distinctive features that are desirable for many real-life problems:

- It provides more information in the interior of a design space.
- It provides a uniform random sampling while treating every design variable as equally important ensuring a uniformly distributed sampling in a given design space.
- The sample size is controllable. The designer, who is usually constrained by the budget, time, or other conditions, determines the sample size.

D. ORTHOGONAL LATIN HYPERCUBE DESIGNS

For the exploratory analysis to be implemented in this thesis, we used Lieutenant Colonel (LTC) Tom Cioppa's recently developed near orthogonal LHDs. These designs were selected as the most appropriate designs for our analysis. The details and a discussion of all the properties of these designs can be seen in Cioppa [Ref.43].

LTC Cioppa developed four sets of 22-factor designs with 129 runs in the first set and 128 runs in the remaining three sets. Note that these runs also correspond to the ranges of the factors. These designs ensure that there is extremely little multi-collinearity across the columns (factors in main effects). In other words, the designs are nearly orthogonal. These designs also have good space-filling properties.

After taking the designs, we modified them by inserting our parameters with their ranges. For some of the parameters (Red number of agents, Max targets per step, combat constraint, and cluster constraint) the range of 129 was illogical. Because of this, we re-scaled them (allowing some duplicates) so that they take reasonable values while 129 runs are evenly distributed within these values. A complete list of these factors and their ranges can be seen in Appendix E.

E. 3^{6-1} FRACTIONAL FACTORIAL DESIGNS

After exploring the data visually and completing the correlation analysis and several regression models, six of the 22 factors noticeably stood out as playing an important role in predicting response variables (MOEs). Upon this discovery, we decided to do a further analysis of these variables. Since the computational capabilities of

factorial designs can handle this size of a problem with only six factors, we used factorial designs. Moreover, since we are primarily interested in main effects and two-term interactions, rather than all possible interactions, fractional factorial designs are used.

However, in factorial experiments, three levels per factor are needed to detect non-linear responses. When the number of factors is large, even a single replicate of a 3^n full factorial design can be expensive and time-consuming [Ref.30]. For instance, one replicate of a 3^6 factorial experiment requires 729 runs. It is often true that at some point higher order interactions tend to become negligible and can be properly disregarded. An appealing strategy in such situations is to make runs for only a fraction of these runs. Provided that care be exercised when choosing factor settings in each run, a great deal of information about factor effects can still be obtained [Ref.30].

At this point we decided to utilize a 3^{6-1} fractional factorial design, (1/3) of the possible 3^6 runs. Our design is a resolution V design. A resolution V design is one that does not confound main effects and two-factor interactions with each other, but does confound two-factor interactions with three-factor and higher interactions. See [Ref.31] for more details on generating a 3^{6-1} fractional factorial design of resolution V.

F. DATA COLLECTION

Today's advances in agent-based models provide the potential to capture some of the adaptability and other key factors inherent in conflict. However, to explore even a single question of interest, a couple of data points are often insufficient based on the landscape of possible outcomes within a given scenario. Thus, in order to gain insight into questions at hand, producing thousands, hundreds of thousands, or even millions of data points is beneficial [Ref.2].

Initially, all three scenarios (see Chapter III for detailed information on these scenarios) were run thousands of times in MANA to explore how appropriately the model represents the real world experience of the author. The secondary intention also was to see the effects of varying parameters on the outcomes of the scenarios. After these preliminary runs, we noticed that two of the original parameters chosen did not affect the outcome as much as the others. As a result of these initial runs, these two parameters

were discarded before the main runs started. Then three scenarios, together with the designs, the model and the parameters, were submitted to powerful computing resources put together to support the Marine Corps Combat Development Command's (MCCDC) processing needs.

As previously stated, we had 22 factors with 129 levels along with three scenarios and four designs. We also wanted to have 100 replications for each run. Putting these together, the total number of runs is

$$3(scenarios) * 513(runs) * 100(replications) = 151{,}900$$

In order to produce the required amount of runs, the Marine Corps' computational capabilities were used. The Marine Corps fielded a cluster of eight Windows NT workstations, each with two processors. This cluster of workstations, termed "Gilgamesh", provides the Marine Corps with an in-house "supercomputer". Housed at Studies and Analysis, the Marine Corps Combat Development Command, Quantico, VA, Gilgamesh provides the ability to run MANA scenarios over 100,000 times in 24 hours [Ref.2].

Gilgamesh completed the 151,900 runs in three days, providing us with the results associated with each set of runs as well as means and quartiles of 100 replications of each run. After further analyzing the factors and determining the most important six factors in Scenario one, a second set of runs was submitted to the supercomputer. This time the base-line scenario, the fractional factorial design, the model, and the parameters of interest were sent over. The total number of runs this time was

$$243(runs) * 100 = 24{,}300$$

These runs, completed in 24 hours, again provided very valuable data to explore.

G. DATA SETS

For the analysis and model development purposes, two data sets are used. The first data set is the result of Near Orthogonal Latin Hypercube Designs on the base-line scenario. This data set contains 51,300 runs and 24 columns. The second data set is the result of the Fractional Factorial Designs on the base-line scenario. This data set includes 24,300 rows and 8 columns. In both data sets two of the columns are the response variables or MOEs and the remaining columns are predictor variables. From now on, the first data set will be referred to as 22-factor Data Set and the second data set will be referred to as 6-factor Data Set.

H. COMPARISONS BETWEEN SCENARIOS

1. Two-Sample Paired Sign Tests

The first two research questions of interest cover the differences between the base-line and two alternative scenarios. The focus is whether a significant difference exists between the scenarios. In essence, we want to compare the base-line scenario with the second and, then, the third scenario. One easy way of making these comparisons is to use the sign test. The sign test often involves the use of matched pairs. It is designed to test a hypothesis that one random variable in a pair tends to be larger than the other random variable in the pair. Specifically, if, for $i = 1, 2, \ldots, n$, X_i and Y_i are paired random variables with $D_i = X_i - Y_i$, the sign test tests the hypothesis that the median of D_i, m_D, is equal to zero. The test statistic *(T)* used is the number of positive differences. If the null hypothesis is true, then the numbers of positive and negative differences should be approximately the same, i.e., T should be close to $n/2$. In this study, the Analyse-It software for Excel is used to perform the sign tests.

In order to construct our test statistics, we now need to identify the null and alternative hypotheses. Under H_o, T is binomially distributed, with parameters n and .5:

- Null hypothesis: $H_0 : m_D = 0$
- Alternative hypothesis: $H_a : m_D \neq 0$

- Level of significance: $\alpha = 0.05$

One way to report the result of a hypothesis test is to simply say whether the null hypothesis was rejected at a specified level of significance. A more informative way to represent this notion of rejection or acceptance is to give the *P*-value. The *P*-value is the smallest level of significance at which H_o would be rejected when a specified test is used on a given data set. Once the *P*-value has been determined, the conclusion at any particular level α results when comparing the *P*-value to α [Ref.30]:

- $P-\text{value} \le \alpha \Rightarrow$ Reject H_o at level α .
- $P-\text{value} > \alpha \Rightarrow$ Do not reject H_o at level α .

I. STATISTICAL TOOLS USED IN THIS STUDY

1. S-PLUS

S-PLUS is a statistical software package to do data analysis and technical graphing. It provides the analysts with the ability to manipulate the data, to graph it and to perform several statistical tests [Ref.20]. In a typical S-PLUS session, the following can be done:

- Import data from virtually any source.
- View and edit data in a convenient data window.
- Create plots with the click of a button.
- Control the details of graphics and produce output for export to the report document.
- Perform statistical analyses from convenient dialogs in the menu system.
- Run analysis functions one at a time at the command line.
- Create individual functions.

In this study, S-PLUS is used for data visualization, regression tree models and regression models.

2. Clementine

Clementine is a data mining application. Data mining offers a strategic approach to finding useful relationships in large data sets. In contrast to more traditional statistical methods, we do not necessarily need to know what we are looking for when starting. Data can be explored, fitting different models and investigating different relationships, until useful information about it is found. The Clementine interface makes data exploration easy. The interface uses an approach called visual programming. Various nodes in the workspace represent different objects and actions. The user connects the nodes to form streams, which, when executed, let the user visualize relationships and draw conclusions. Streams are like scripts: the user can save them and reuse them with different data files [Ref.23]. In this thesis, Clustering Analysis and Neural Networks are done using Clementine.

3. Netica

Netica is an application for working with Bayesian belief networks and influence diagrams. It contains functions to build, learn, modify, transform, save and read networks, as well as having a powerful inference engine [Ref.21]. Characteristics of Netica are

- It has operations to build and modify networks in memory, as well as to save them to file and read them back.
- It can solve decision networks to find optimal sequences of decisions and conditional plans.
- It can learn probabilistic relations from data.
- It can learn from experience using networks, which are being queried to do probabilistic inference.
- It can reverse individual links (while maintaining the same joint probability distribution) and "sum out" nodes of an influence diagram or belief net.
- It can save individual cases to file, and manipulate files of cases.

- It can perform sampling (discrete simulation) to generate random cases with a probability distribution matching the belief network.
- It has nodes and networks with a "user-settable/readable" field, which can contain a pointer to anything chosen.

Netica is used for exploratory analysis of the data. It provides visual relational representation of the data. Additionally, it is used to construct a predictive model and answer research questions.

4. Multiple Additive Regression Trees (MART)

Multiple Additive Regression Trees, MART, is a methodology for predictive data mining [Ref.25] that is usable with S-PLUS as an enhancement to it, especially in regression tree methods. In this study, MART is used to develop a model for 22-factor Data Set together with regular regression tree techniques, which are built in S-PLUS.

5. GGobi

The GGobi software is a data visualization system with “state-of-the-art” interactive and dynamic methods for the manipulation of views of data [Ref.24]. It has the same graphical functionality whether it is running standalone or embedded in other software. That functionality includes 2-D displays of projections of points and edges in high-dimensional spaces, as well as scatter-plot matrices, parallel coordinate and time series plots. Projection tools include average shifted histograms of single variables, plots of pairs of variables, and grand tours of multiple variables. Views of the data can be reshaped. Several displays can be open simultaneously and linked for labeling and brushing. Missing data are accommodated and their patterns can be examined [Ref.24]. In our analysis, GGobi is used extensively for visualizing both data sets.

J. DATA VISUALIZATION

Visual methods have a special place in data exploration because of the power of the human eye/brain to detect structures. These methods are important because they are ideal for sifting through data to find unexpected relationships [Ref.29].

1. Tools for Displaying Single Variables

One of the most basic displays for viewing a single response variable is the histogram, showing the number of values of the variable that lie in consecutive intervals. In large data sets, especially, the histogram can reveal real aspects of the underlying distribution of that single variable of interest [Ref.29]. In our analysis, S-Plus generated histograms are used.

It is often desirable to display different groups of results on a single response variable separately so that the groups may be compared. In such cases, a useful alternative display is the boxplot. Boxplots contain a box with the bulk of the data, for example, the interval between the first and third quartiles. A line across this box indicates the median of the data [Ref.29]. In this analysis, S-Plus generated boxplots of the data are used to demonstrate the difference between scenarios in terms of MOEs.

2. Tools for Displaying Relationships between Two Variables

Scatter and line plots are the most basic kinds of plots for displaying data. They can be used to plot a single column of data or to plot one data column against another.

The scatter plot is a standard tool for displaying two variables at a time. Scatter plots are similar to line graphs in that they use horizontal and vertical axes to plot data points. However, they have a very specific purpose. Scatter plots show how much one variable is affected by another variable. In other words, they represent how much two variables are correlated with each other [Ref.29].

3. Tools for Displaying More Than Two Variables

Scatter plots can also be combined in multiple plots per page to help understand higher-level structure in data sets with more than two variables. This set of plots is called scatter plot matrix. One limitation of the scatter plot matrix is that it cannot show interaction effects with another variable. At this point, conditioning plots, also known as Trellis plots, are very useful. A Trellis plot is a plot of two variables that are conditional on the value of a third variable, called as conditioning variable. Bar charts and surface plots are also used for visualization of the data sets. Bar charts represent a requested

statistic based on the values of one or more variables. They are useful for displaying exact magnitudes and emphasizing differences among the charted values. Surface plots are three-dimensional surfaces reflecting the value of a variable over different values of two other variables.

K. MODEL DEVELOPMENT

1. Descriptive and Predictive Modeling

Before elaborating further on various analysis tools and techniques used in this study, we want to clarify what the terms model, pattern, descriptive modeling and predictive modeling mean. A model is a high-level description, summarizing a large collection of data and describing its important features. Often a model is global in the sense that it applies to all points in the measurement space. In contrast, a pattern is a local description, applying to some subset of the measurement space, perhaps showing how just a few data points behave or characterizing some persistent but unusual structure within the data [Ref.29].

A descriptive model presents, in convenient form, the main features of the data. It is essentially a summary of the data, permitting the study of the most important aspects of the data while perhaps providing the patterns associated with the variables [Ref.29].

On the other hand, the purpose of a predictive model is to estimate one or more dependent variables as accurately as possible from a set of independent variables. One way to do this is to fit the model to data. Data used to build such models usually come from closely monitored training sets. The variable to be predicted and its covariates are carefully measured to build a training data set. After training, the fitted model must be validated with a test set, a set of data outside the training set that gives a way to measure the model's ability to generalize what it has learned [Ref.29].

In this study, the focus is on understanding the data and developing a plausible model than can be used to predict future data. In essence, both descriptive and predictive modeling techniques are required. In order to summarize the data and study the most

important characteristics of it, Cluster Analysis is used. This method provided us the interesting patterns related to our response variables (MOEs).

For predictive modeling a variety of analysis methods are used. These methods are Bayesian Networks, Multi Adaptive Regression Trees (MART), Regression Trees, Neural Networks, and classical Multiple Regression Techniques. For the 22-factor data set all the modeling techniques are used except Multiple Regression. For the 6-factor data set all the techniques are applied. In all of the predictive modeling methods, the data sets are split into training and test sets. For these purposes, random sampling algorithms of statistical packages are used.

2. Random Sampling

In random sampling, each item or element of the population has an equal chance of being chosen at each draw. A sample is random if the method for obtaining the sample meets the criterion of randomness. The statistical packages, S-Plus and Clementine, used in this study, have their own algorithms for random sampling. In S-Plus, random sampling can be done automatically by using a built in random sampling function. In Clementine this can be accomplished by using several nodes. For our analysis purposes, we selected 80 percent of the data as the training set and the remaining 20 percent as the test set. These sets are created for all techniques except Neural Networks. Neural Networks require smaller data sets to implement. After many trials it was discovered that the largest data size with 22 predictors that Clementine's Neural Networks can handle is around 10,000. Therefore, instead of 80 percent, 20 percent of the 22-factor data set was randomly chosen as the training set and another 20 percent was chosen as the test set.

3. Cluster Analysis (Clementine)

Cluster analysis is basically decomposing or partitioning a data set into groups so that the points in one group are similar to each other and are as different as possible from the points in other groups. The purpose of doing this analysis is to see whether the data is composed of natural subclasses [Ref.29]. We expect to discover something about the nature of the data. Obviously, the clustering methods focus on the notion of distance. As

far as these techniques are concerned, the concept of distance is more fundamental than the coordinates of the points. In principle, to carry out a cluster analysis all that should be known is the set of inter-point distances [Ref.29]. For further details of Cluster Analysis used in this thesis see [Ref.23].

4. Multiple Additive Regression Trees (MART/S-Plus)

A particular methodology for trying to solve prediction problems is MART. In predictive modeling, the goal is to use the data to produce an accurate mapping over the response variable. The notion of accuracy depends on the type of the response variable y in terms of the set values it can assume. If y assumes numeric values, the problem is known as regression and lack of accuracy is defined in terms of a distance measure between the predicted value and the unknown true value of y [Ref.25]. Two common measures of inaccuracy are average absolute error

$$AAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}$$

or root mean squared error

$$RSME = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}$$

In these equations, y_i represents actual values and y_i represents fitted values of the response variables. The difference between them $y_i - \hat{y}_i$ is called residual or error.

Besides accuracy, the other primary goal of MART is robustness. It tends to be resistant against moderate to heavy contamination by bad measurements (outliers) of the predictors and/or responses, missing values, and to the inclusion of potentially large numbers of irrelevant predictor variables that have little or no effect on the response [Ref.25]. Internally, MART randomly partitions the input data into a training data set and a complement test set. Another advantage of MART is that it shows the estimated relative importance of each predictor variable in predicting the response based on the

current model. Additionally, MART provides the dependence of the model on each factor by identifying the most relevant predictor variables. See [Ref.25] for more details on MART and its applications.

5. Netica Bayesian Networks (Netica)

The graphical models that we have used to represent dependencies between data and parameters associated with the sampling model can be further used to describe directed associations among sets of variables. When used this way, such models are known as Bayesian Belief Networks (BBN) [Ref.22]. A belief network, also known as a Bayesian network or probabilistic causal network, captures our assumed relations (which may be uncertain, or imprecise) between a set of variables that are relevant to some other variables. They might be relevant because we will be able to observe them, because we need to know their value to take some action or report some result, or because they are intermediate or internal variables that help to express the relationships between the other variables [Ref.21]. For further discussion and details on BBNs see [Ref.21] and [Ref.22].

6. Neural Networks (Clementine)

Neural networks are one of a class of highly parameterized statistical models that have attracted considerable attention in recent years. The fact that Neural Networks are highly parameterized makes them very flexible, so that they can accurately model relatively small irregularities in functions. On the other hand, such flexibility means there is a serious danger of over-fitting [Ref.29]. See [Ref.29] and [Ref.23] for more details on Neural Networks and its applications in Clementine.

7. Regression Trees (S-Plus)

Tree-based models provide an alternative to linear and additive models for regression problems and to linear and additive logistic models for classification problems. Tree models are fit by successively splitting the data to form homogeneous subsets. The result is a hierarchical tree of decision rules useful for prediction or classification [Ref.20]. The basic principle of tree models is to partition the space spanned by the input

variables to maximize a score of class purity, meaning (roughly depending on the particular score chosen) that the majority of the points in each cell of the partition belong to one class. Tree models have many attractive properties while being easy to understand and explain. They can handle the mixed variables (continuous and integer, for instance) with ease since trees partition the space using binary tests, and they can predict the class value for a new case very quickly. They are, also, very flexible, so that they can provide a powerful predictive tool [Ref.29]. See [Ref.20] and [Ref.29] for more details on Regression Tree models.

8. Multiple Regression Techniques

Regression is the standard technique for assessing how various predictors relate to a response. These are the regression techniques considered in this study:

- **Linear regression**: Predicting a continuous response as a linear function of predictors using a least-squares fitting criterion.
- **Generalized additive models:** Predicting a general response as a sum of nonparametric smooth univariate functions of the predictors.

a. Linear Regression

Linear regression is used to describe the effect of continuous or categorical variables upon a continuous response. It is by far the most common data-based model building procedure. The linear regression model assumes that the response is obtained by taking a specific linear combination of the predictors and adding random variation (error). The error is assumed to have a Gaussian (normal) distribution with constant variance and to be independent of the predictor values [Ref.29]. Linear regression models can be effectively used to measure the effects of interactions as well as main effects. For further information on Linear Regression models see [Ref.29].

b. Generalized Additive Models (GAMs)

Generalized additive models extend linear models and generalized linear models by flexibly modeling additive nonlinear relationships between the predictors and the response. Whereas linear models assume that the response is linear in each predictor, additive models assume only that the response is affected by each predictor in a smooth way. The response is modeled as a sum of smooth functions in the predictors, where the smooth functions are estimated automatically using smoothers. The GAM suggests a possible curved relationship, which may more accurately reflect the progress of the condition [Ref.20].

GAMs are useful where

- The goal is to obtain a final fit or to explore what types of variable transformations might be appropriate for use in a standard linear model;
- The relationship between the variables is expected to be of a complex form, not easily fitted by standard linear or non-linear models;
- No *a priori* reason for using a particular model exist; and
- The data should suggest the appropriate functional form.

See [Ref.35] for more details on GAMs. In this chapter, the MOEs, the designs and the statistical techniques used in the analysis were explained in detail. Data collection and data visualization methods were mentioned. In the next chapter, the results of the analysis will be presented and explained.

V. RESULTS

“The skillful leader subdues the enemy's troops without any fighting; he captures their cities without laying siege to them; he overthrows their kingdom without lengthy operations in the field. With his forces intact he disputes the mastery of the empire.”

Sun-Tzu

A. CHAPTER OVERVIEW

This chapter explains the results of the statistical methods applied to the two different data sets (Latin Hypercube Design and Fractional Factorial Design) at hand. Each measure of effectiveness is examined while the data sets are explored to determine the significance of the parameters as well as the differences between scenarios. Sign tests are used to determine whether there is a significant difference between scenarios in terms of MOEs. Various visualization tools from several statistical packages (S-Plus, Ggobi, Netica) are used to build a better understanding about relationships between different variables. Models are developed using some statistical tools (S-Plus, Netica, MART, Clementine) for descriptive and predictive analysis of the data sets. Finally, the research questions discussed earlier are examined using these results.

B. TWO-SAMPLE PAIRED SIGN TESTS

Frequently the objective of a test procedure is to discover if there is a significant difference between population parameters of interest rather than to estimate these parameters [Ref.30]. A non-parametric method of accomplishing this statistical inference is called the sign test. In this study, a series of sign tests are implemented to determine if there is a significant difference between the base-line and two alternative scenarios. The averages of each 100 replications are taken from the data sets of each scenario. The random differences D_i, are then calculated for each of the 513 input combinations. If there is no difference between the scenarios, then we expect the median of D_i, m_D, to be zero. We use the sign test to test this.

- Null hypothesis: $H_0 : m_D = 0$
- Alternative hypothesis: $H_a : m_D \neq 0$
- Level of significance: $a = 0.05$

Under H_o, T is binomially distributed, with parameters n and .5.

1. Sign Tests for Difference between Red Proportions

Boxplots in Figure 7 show the differences between Scenarios. In this plot, the y-axis corresponds to the values of the differences while the x-axis represents the differences between scenarios. The boxplot on the left, which is assigned to 1, corresponds to the difference between Scenario one and Scenario two. The boxplot on the right corresponds to the difference between Scenario one and Scenario three. These boxplots suggest that no significant difference exists between Scenario one and two whereas a significant difference exists between Scenario one and three in terms of the proportion of Red killed. In other words, different formation of the Blue vehicles/infantry does not have an effect on the outcome of the battle while three-lane formation of Red does have a significant effect on the outcome in terms of Red casualties. Three-lane formation increases the proportion of Red killed.

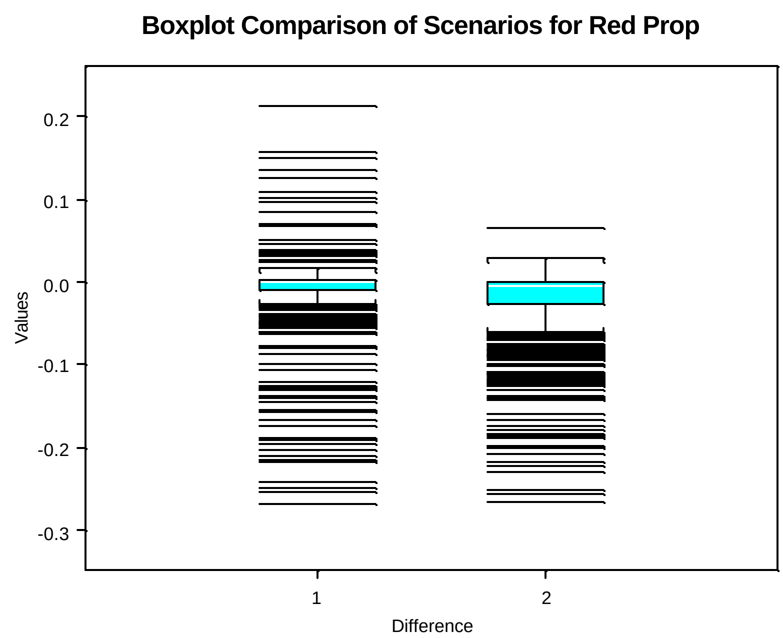

Figure 7. Boxplot for the comparison of the scenarios (MOE1)

a. Scenario one and Scenario two

Let's assume m_1 is the number of negative differences, that is the proportion of Red killed in Scenario one is greater than the proportion of Red killed in Scenario two, m_2 is the number of positive differences, and m_3 is the number of ties. Finally, n is the total number of positive and negative observations discarding the ties.

$m_1 = 235$ T = $m_2 = 263$ $m_3 = 15$ $n = 498$

$P - value = 0.226$

Since 263 ≈ 235 and the P-value = 0.226, we do not reject the null hypothesis. We conclude that there is no significant difference between these two scenarios in terms of Red casualties. In other words, the different formation of the Blue vehicles and infantries does not have a significant effect on the number of Red agents killed.

b. Scenario one and Scenario three

$m_1 = 120 \qquad m_2 = 385 \qquad m_3 = 8 \qquad n = 505$

$P - value = 0$

Since 385 >> 120 and the P-value = 0, we reject the null hypothesis. We conclude that there is a significant difference between these two scenarios. In other words, the third infiltration team of the Red agents has a significant effect on the number of Red agents killed; this formation increases the number of Red casualties.

2. Sign Tests for Difference between Blue Proportions

Boxplots in Figure 8 show the differences between Scenarios based on the proportion of Blue killed. The boxplot on the left, which is assigned to 1, corresponds to the difference between Scenario one and Scenario two. The boxplot on the right corresponds to the difference between Scenario one and Scenario three. These boxplots suggest that a significant difference exists between Scenario one and two whereas no significant difference exists between Scenario one and three in terms of the proportion of Blue killed. In other words, different formation of the Blue vehicles/infantry does have a significant effect on the number of Blue killed while three-lane formation of Red does not have an effect on the outcome in terms of Blue casualties. The different formation of the Blue vehicles/infantry decreases the proportion of Blue killed.

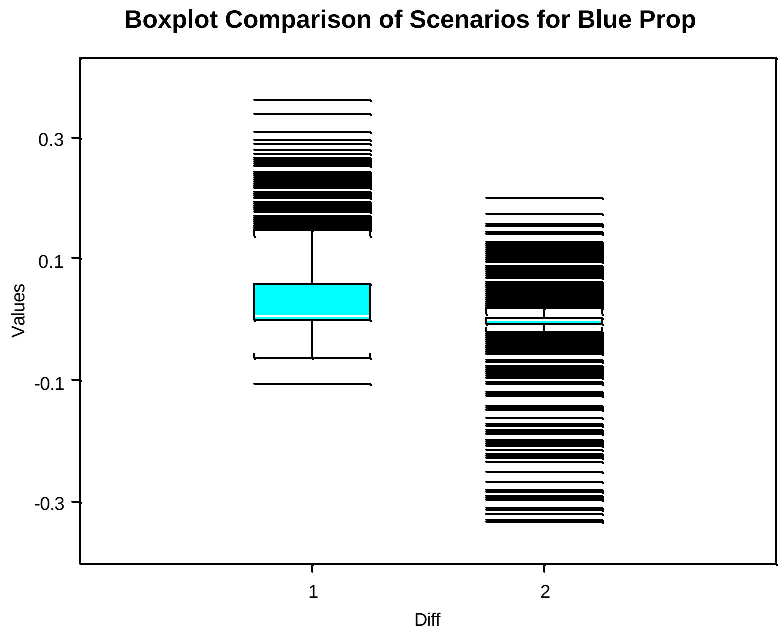

Figure 8. Boxplot for the Comparison of the Scenarios (MOE2)

a. Scenario one and Scenario two

$m_1 = 81$ $m_2 = 381$ $m_3 = 51$ $n = 462$

$P - value = 0$

Since 381 >> 81 and the P-value = 0, we reject the null hypothesis. We conclude that there is a significant difference between these two scenarios. In other words, the different formation of the Blue vehicles and infantries has a significant effect on the number of Blue agents killed; this formation reduces the number of Blue casualties.

b. Scenario one and Scenario three

$m_1 = 242 \qquad m_2 = 222 \qquad m_3 = 49 \qquad n = 464$

$P - value = 0.378$

Since 242 ≈ 222 and the P-value = 0.378, we do not reject the null hypothesis. We conclude that there is no significant difference between these two scenarios. In other words, the third infiltration team does not have a significant effect on the number of Blue agents killed.

C. CLUSTER ANALYSIS IN CLEMENTINE

Cluster Analysis is a good way of describing and searching for patterns in a data set. For this reason, the clustering method in Clementine is used to examine the relationship between two MOEs, the proportion of Red killed and the proportion of Blue killed. The goal is to find out which results lead to other results on the enemy side. Before using the clustering model, the values of MOEs in data sets are transformed from numerical values into categorical values. As previously given, our MOEs take continuous values between 0 and 1. We split these values into three categories of low, medium, and high, by defining some ranges. For both MOEs, the values from 0 to 0.25 are assigned a value of 0 (low), the values from 0.25 to 0.75 are assigned a value of 1 (medium), and the values bigger than 0.75 are assigned a value of 2 (high).

1. The 22-Factor Data Set

a. Red Killed

For the Red side, the proportion of Red killed takes a high value almost 85 percent of the time. That is Red gets good results in 15 percent of the runs. In our further analysis, we will concentrate on the results where Blue has high or medium casualties and Red has medium or low casualties. We will attempt to find the factors that are associated with these results. The histogram of MOE1 (Figure 9) supports the results of cluster analysis.

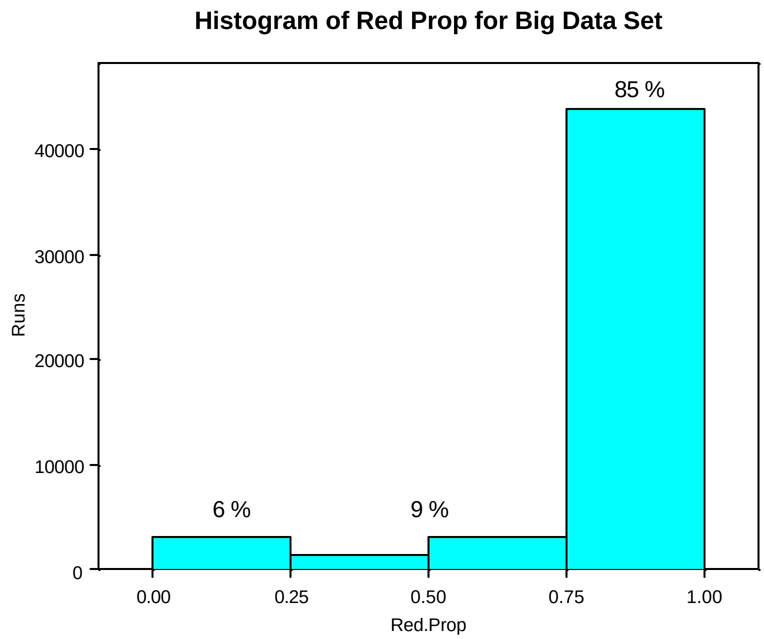

Figure 9. Histogram of MOE1 for the 22-factor Data Set

b. Blue Killed

For MOE2 the results are in favor of Blue for most of the runs, because approximately 80 percent of the time Blue takes low casualties. The clustering analysis suggests that we should focus on the cases where Blue has medium or high casualties, which are 19 percent and 1 percent respectively. The histogram below also supports this analysis. The runs where the blue casualties are medium are especially worth looking at closely since they occur in almost one of five cases.

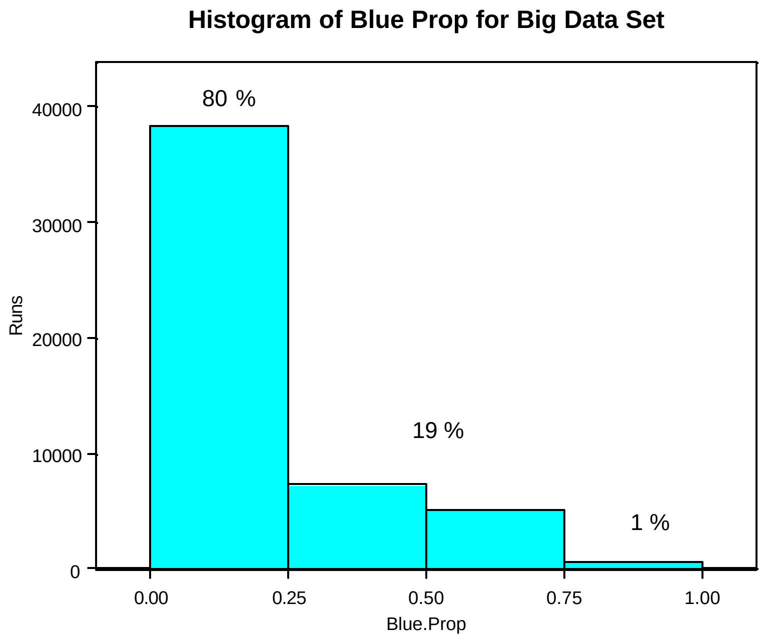

Figure 10. Histogram of MOE2 for the 22-factor Data Set

c. Blue and Red Killed

The primary goal of cluster analysis is to find the patterns between the two MOEs. The data here is divided into five clusters based on the distances between the values of the MOEs. In Cluster-1, 24 percent of the data, the Red side performs well by inflicting medium casualties on the Blue side while taking high casualties only 50 percent of the time. In Cluster-2, 2 percent of the data, both sides take only low casualties. In Cluster-3, 1 percent of the data, the Red team takes medium casualties while the Blue team takes low. In Cluster-4, 1 percent of the data, the Blue team takes high casualties while the Red team also takes high casualties 74 percent of the time. Finally, in Cluster-5, 72 percent of the data, the Red force takes high casualties while the Blue force takes low casualties.

Figure 11 shows these clusters and the relationships between MOEs. This is a Trellis plot of our MOEs conditioned on five different clusters. The y-axis represents the proportion of Red killed while the x-axis represents the proportion of Blue killed. In this Trellis Plot, Cluster-1 and Cluster-4 suggest that Red is successful by inflicting

medium and high Blue casualties while suffering less high losses. By looking at the parameter settings associated with Cluster-1 and Cluster-4 some patterns and relationships can be detected between predictor and response variables.

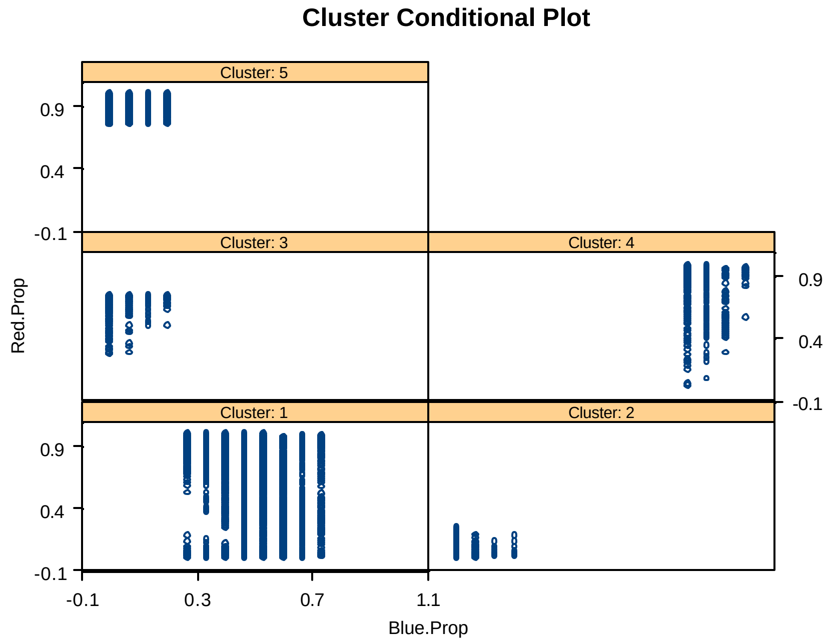

Figure 11. Cluster Trellis Plot for the 22-factor Data Set

Figure 12 gives a relational representation of how clusters are formed. In this plot Cluster-1 is the highest point in the Cluster column and demonstrates an interesting pattern of good results for the Red side. Also, Cluster-3 and Cluster-4 show important patterns for the analysis. In these clusters, Red scores good results by taking low casualties and inflicting high casualties on Blue.

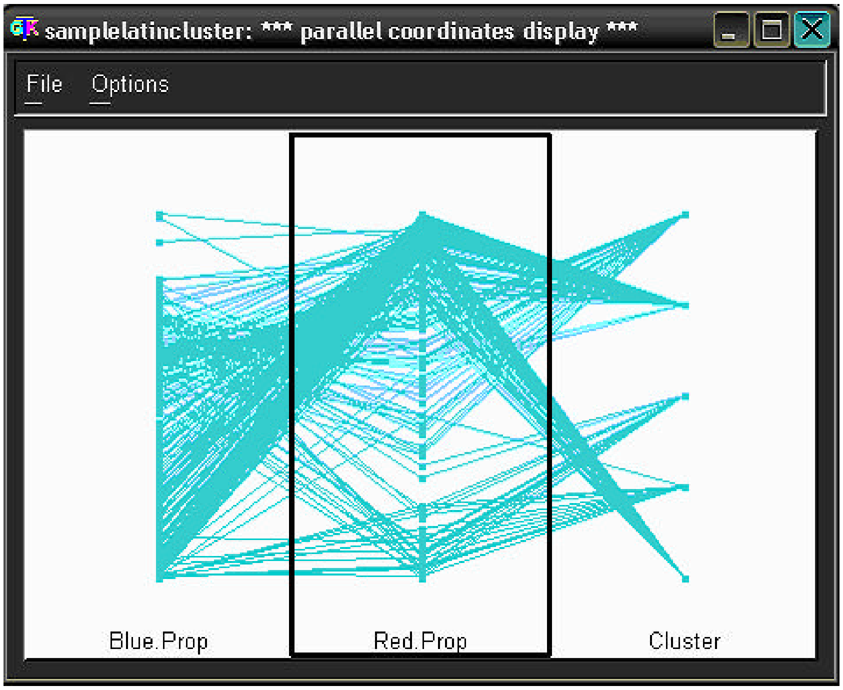

Figure 12. Parallel Coordinates Plot of Clusters for the 22-factor Data Set

2. The 6-Factor Data Set

a. Red Killed

Cluster analysis is also applied to the 6-factor data set. The values of MOE1 in this data set reveal an interesting pattern. The medium values for the proportion of Red agents killed are almost never observed while the low and high values appear frequently depending on the settings of the predictor variables. After a close look at the data, we detected that MOE1 takes values either below 0.2 or above 0.75. The histogram below shows this situation more clearly.

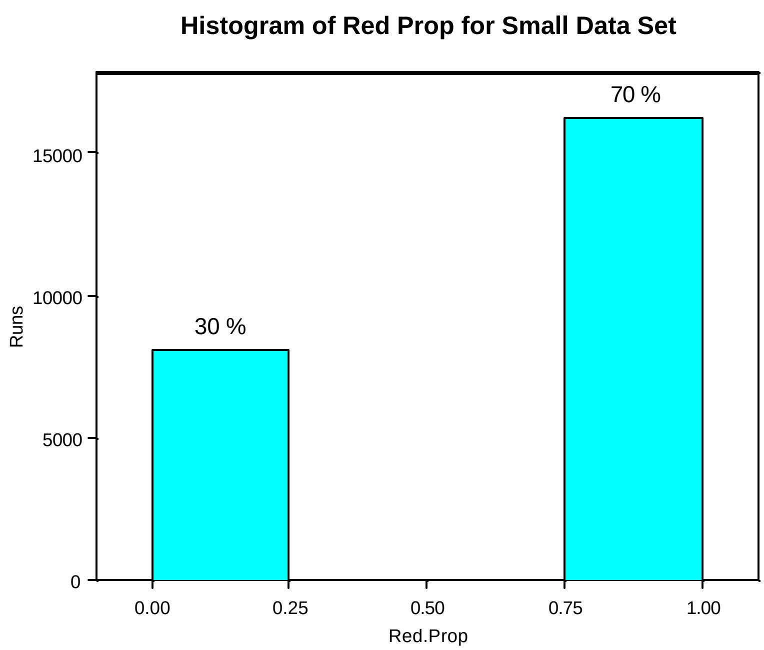

Figure 13. Histogram of MOE1 for the 6-factor Data Set

b. Blue Killed

Clustering of MOE2 provides three clusters associated with low, medium and high values of Blue killed. These clusters also suggest that 75 percent of the time Blue takes low casualties. However, this time Blue takes medium casualties more frequently, 24 percent of the cases, than the 22-factor data set. This observation can be a result of a specific variable or variables in the 6-factor data set. The histogram below also suggests that the parameters associated with the proportion of Blue killed values of 0.5 to 0.75 should be explored.

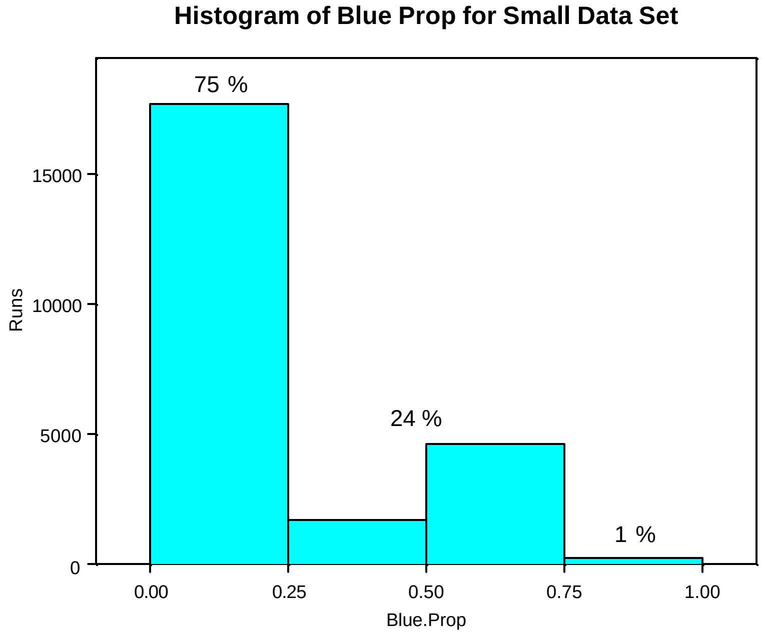

Figure 14. Histogram of MOE2 for the 6-factor Data Set

c. Blue and Red Killed

When a cluster analysis is done on both MOEs, more interesting results and patterns are discovered. For the 6-factor data set, five clusters are formed representing the relationships between MOE1 and MOE2. According to Cluster-1, 69 percent of the data, the Blue side does not lose many agents while inflicting high Red casualties. Cluster-2, 20 percent of the data, suggests that the Red team usually takes low casualties while causing the Blue team high losses. Cluster-3 and Cluster-4, only 1 percent of the data, says that the Red team suffers low or medium casualties while imposing high losses on the Blue team. Finally, Cluster-5, 10 percent of the data, shows that both sides suffer low casualties. Figure 15 illustrates these clusters. Cluster-2 and Cluster-4 seem to be very advantageous for the Red side with low friendly and high enemy losses. The parameter settings associated with these clusters may reveal interesting results on the effects of specific predictor variables.

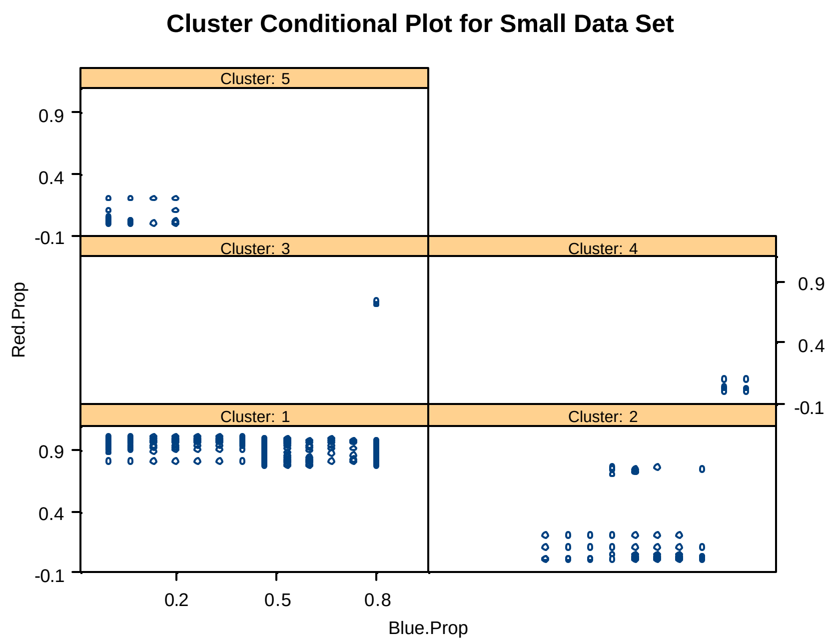

Figure 15. Cluster Trellis Plot for the 6-factor Data Set

Figure 16 demonstrates a relational structure of how clusters are formed. In this plot, Cluster-2 (second point from the top in Cluster column) shows a critical relationship between MOEs where the proportion of Red killed is usually low while the proportion of Blue killed is high. This can be identified by looking at the lines coming from the Blue Prop and going into Cluster-2 through the Red Prop. Higher Blue Prop values match with lower Red Prop values to form Cluster-2. The other importance of this cluster is that it contains a big portion of the data, 20 percent.

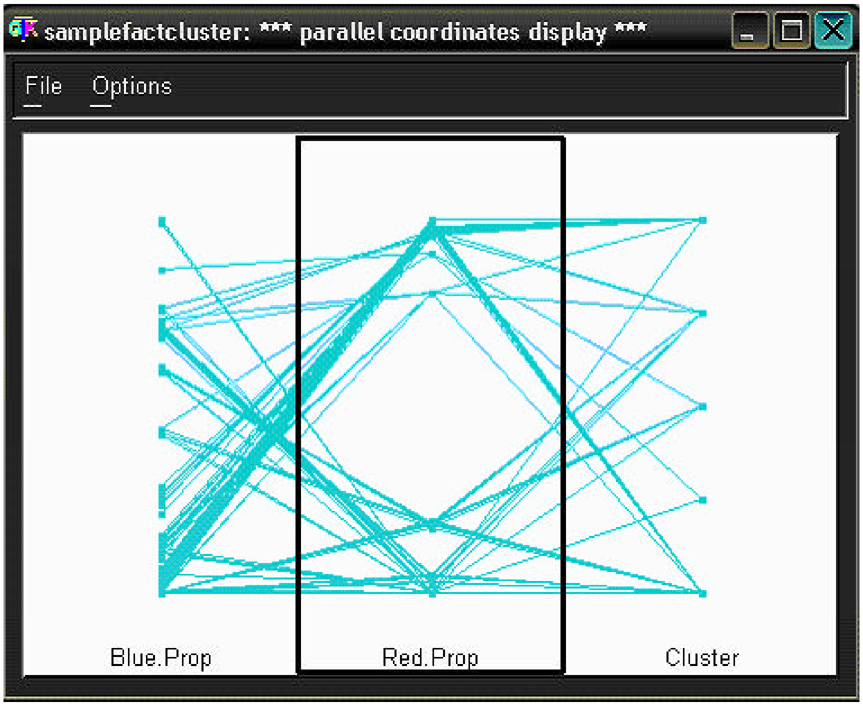

Figure 16. Parallel Plot of Clusters for the 6-factor Data Set

D. NEURAL NETWORK MODEL IN CLEMENTINE

The first model built for predictive modeling is the Neural Network model. For constructing this model Clementine is used. Neural Networks adopt a basic idea of transforming a linear combination of the predictor variables via nonlinear transformations using multiple layers. The model contains three layers, representing inputs, hidden layer, and outputs. The Hidden layer is where all the linear combinations of the non-linear transformations of the predictor variables take place.

1. The 22-Factor Data Set

As previously given, the 22-factor data set has 51,300 rows, which is substantial. Since Clementine's Neural Network does not work with huge data sets, 10 percent of the data is taken as a training set by using a random sampling (without replacement) algorithm in Clementine. Another 10 percent of the remaining data is randomly assigned to be the test set. Since we have 22 predictors, the input layer has 22 nodes. The Hidden layer has eight nodes and the output layer has one.

a. Red Killed

After training the data with the response variable, MOE1, the model provides us a mean absolute error of 0.0328 on the training set. It also gives us the relative importance of the variables. According to this model, the most important six variables for predicting Red casualties are the Red Stealth, the Red Movement, the Red Propensity to Enemy (w2), the Blue Infantry Firing, the Red Number of Agents, and the Red SSKP. Since there is always a danger of over-fitting in Neural Network Models, test data is used to see the predictive power of this model. The mean absolute error of the model on the test set is 0.0334, which is about the same as the training set error.

b. Blue Killed

The same type of Neural Network model for MOE2 results in a mean absolute error of 0.0477 on the training set. According to this model, the most important six variables for predicting Blue casualties are the Red Stealth, the Red Number of Agents, the Red SSKP, the Recon Firing, the Recon Stealth, and the Red Combat. Here note that the Red Stealth, the Red Number of Agents, and the Red SSKP parameters are involved in the previous model of MOE1. The mean absolute error of the model on the test set is 0.0483, which is almost the same as the error of the training set.

2. The 6-factor Data Set

The 6-factor data set consists of 24,300 runs with 8 variables. In order to use the whole data, 80 percent of the data is randomly chosen to be the training set while the remaining 20 percent is selected to be the test set. In this Neural Network Model, the input layer has six nodes, the hidden layer has eight nodes and the output layer has one node.

a. Red Killed

When we consider MOE1, the model predicts the output with a mean absolute error of 0.0128 in the training set, which is smaller than the error obtained from the Neural Network model of the 22-factor data set. The most important three variables

are the Red Stealth, the Recon Stealth, and the Number of Red Agents. The prediction of the test set with this model gives us a mean absolute error of 0.013.

b. Blue Killed

For MOE2, the mean absolute error on the training set is 0.04, which is also smaller than the error obtained from the model of the 22-factor data set. The most important three variables are the Red Stealth, the Red SSKP, and the Number of Red Agents. Finally, the mean absolute error for the test set is 0.105.

In conclusion, the Neural Network Model on the 22-factor data set gives us very small errors, which are also consistent with each other in the training and test sets. However, this model is very complicated with 22 input variables, especially when considering all the linear combinations of these variables in the hidden layer. The model used for the 6-factor data set gives good predictive power for MOE1, but it does not do well on MOE2.

E. THE REGRESSION TREE MODEL FOR THE 22-FACTOR DATA SET

For the 22-factor data set, Regression Tree models supported by MART are used. The plots below showing the relative importance of the variables in the tree models are taken from MART and are incorporated in Regression Tree models of S-Plus.

1. Red Killed

In Figure 17, the y-axis represents the most important predictor variables and the x-axis represents the importance of these variables in the tree model. One important observation in this plot is that the Red Movement parameter turns out to be the second most important parameter for MOE1. Another interesting point is that the Red personality parameters do not have much influence on the proportion of Red killed.

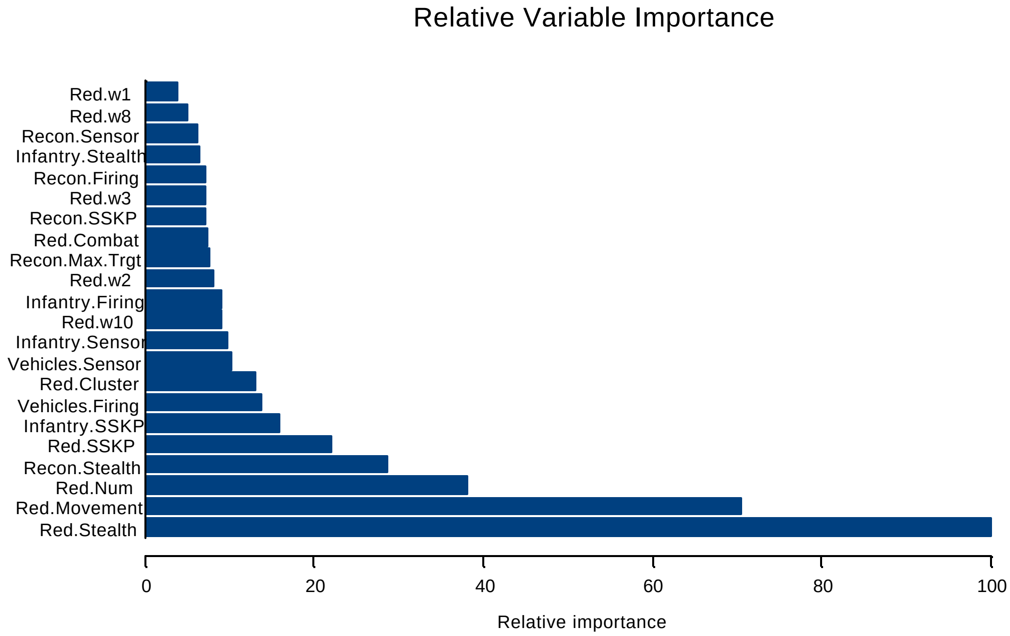

Figure 17. Relative Importance of the Variables for MOE1 in MART

```
Variables actually used in tree construction:
"Red.Stealth" "Red.Movement" "Red.Num"
"Recon.Stealth"
Number of terminal nodes:  13
Residual mean deviance:  0.006004 = 240.1 / 39990
```

Table 1. Regression Tree Model of MOE1 for the 22-factor Data Set

In this first tree model for MOE1, Table 1 shows we have 13 nodes, which means 13 different results based on the values of parent variables. Residual mean deviance is the sum of the squared values of residuals divided by total number of observations. Residuals are simply the differences between actual and predicted values. Many tree models are searched to find the simplest and the most efficient model with the lowest mean deviance. The residual mean deviance in this model is 0.006. Comparing the variables used in this model with the most important four variables of the MART model indicates that these two models are fairly consistent with each other. The residual mean deviance for the test set is 0.0064. The tree model for MOE1 is below.

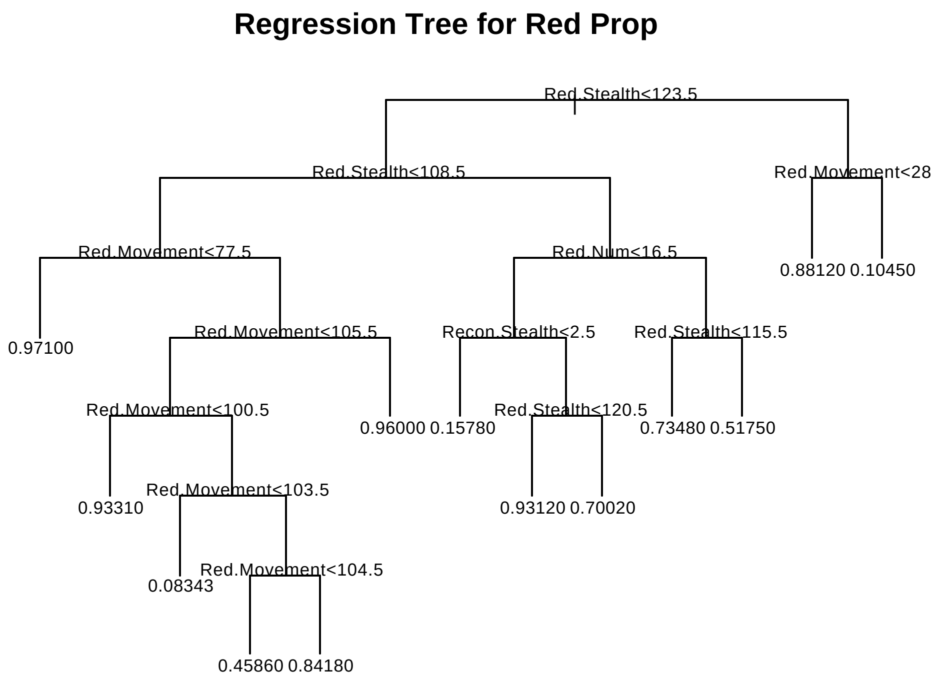

Figure 18. Regression Tree Model of MOE1 for the 22-factor Data Set

These regression tree models seem quite reasonable with strong predictive power. They are also easy to interpret with a good visual representation of the model. Reading the model is best illustrated by taking the first model in Figure 18 and following the reasoning step by step. If the Red Stealth is greater than 123.5 and the Red Movement is less than 28 then the Red side takes around 88 percent casualties. In other words given these conditions, no matter what values other parameters take the Red team takes high casualties.

2. Blue Killed

Figure 19 shows the important variables for MOE2. When looking at the most important six parameters, four of them are the same as the most important parameters of the model of MOE1. Another crucial finding is that all the important parameters are associated with Red infiltration and recon teams.

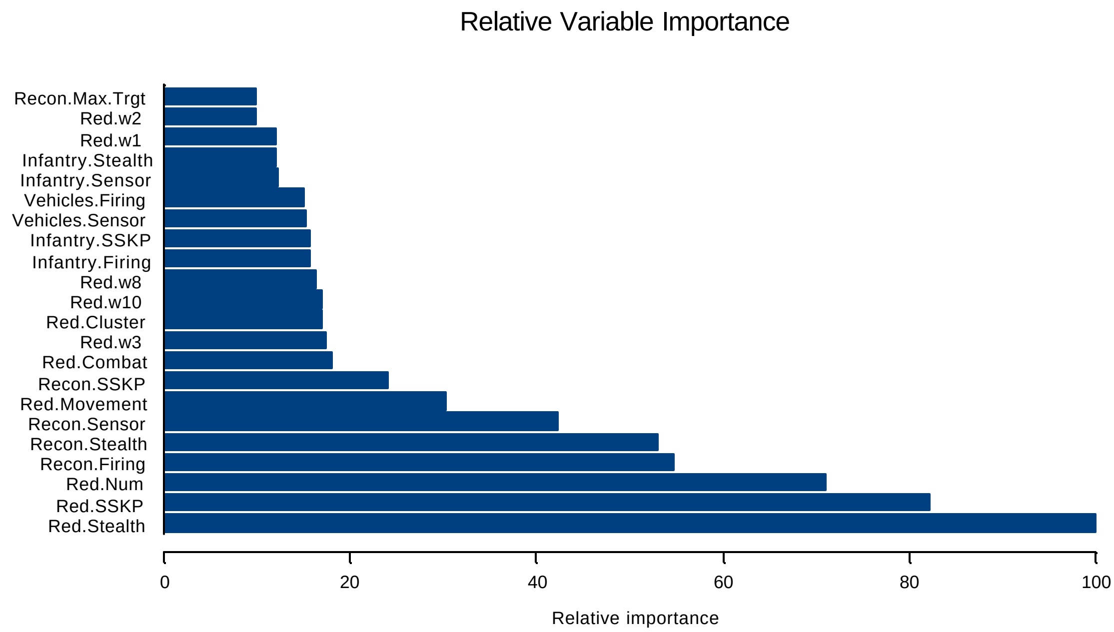

Figure 19. Relative Importance of the Variables for MOE2 in MART

```
Variables actually used in tree construction:
"Red.Stealth"  "Red.Num"  "Recon.Stealth"
"Recon.Firing"  "Recon.Sensor"  "Red.SSKP"  "Red.w1"
Number of terminal nodes:  13
Residual mean deviance:  0.0165 = 659.9 / 39990
```

Table 2. Regression Tree Model of MOE2 for the 22-factor Data Set

Table 2 illustrates that this model contains 13 nodes or leaves. When the variables used in the tree model are compared with the important variables of the MART model, they are fairly consistent. To measure the predictive power of this model, the test data set is used again. The residual mean deviance of the tree model for the test set is 0.016. Finally, Figure 20 is our tree model of the proportion of Blue killed for the 22-factor data set.

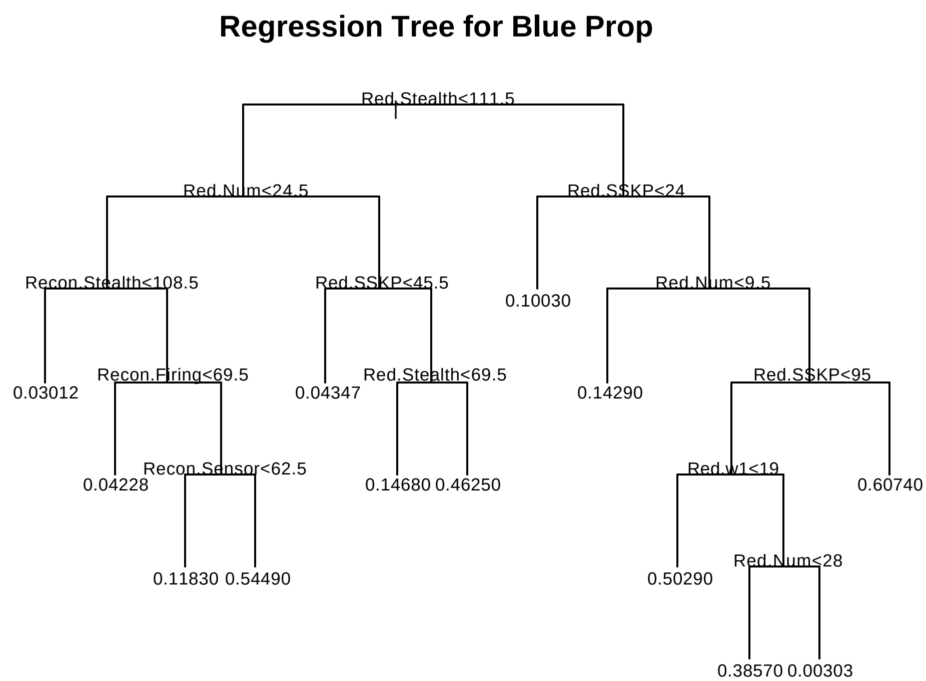

Figure 20. Regression Tree Model of MOE2 for the 22-factor Data Set

In Figure 20, if the Red Stealth is less than 111.5, the Number of Red Agents is less than 24.5 and the Recon Stealth is less than 108.5, Blue takes very low casualties. In other words, given these conditions no matter what values other parameters take, the Red team cannot cause any trouble for the Blue team. The tree models developed for MOEs provide us very strong predictive power with simple and good visual representations. One important observation is that the Red parameters are decisively important for both MOEs.

F. NETICA BAYESIAN NETWORK MODELS

Another model developed to predict the MOEs based on the values of predictor variables is the Bayesian Belief Network. A belief network, also known as a Bayesian network or probabilistic causal network, captures our assumed relations (which may be uncertain, or imprecise) between a set of variables or nodes relevant to some other variables. Probabilistic relations are provided for each node, which express the

probability of that node having different values conditioned on the values of its parent nodes. Many preliminary models with different predictor variables and different model sizes are tested to locate a simple and an effective model with good predictive power. For both 22-factor and 6-factor data sets, all of the variables are transformed to categorical variables with three levels, namely low, medium and high. This is accomplished by partitioning the values of the variables close to thirds. The ranges of all variables can be seen in Appendix E. The data sets also split into training and test sets. The training set consists of a randomly selected 80 percent of the data. The test set includes the remaining portion of the data.

1. The 22-factor Data Set

After many trials of different networks with various model sizes, we realized that an eight-node model for the 22-factor data set is simple and effective enough to meet our demands.

a. Red Killed

Among the several models with eight nodes, we found the best one for MOE1 with the lowest prediction error rate. Figure 21 demonstrates the nodes and the relational arcs between them. The nodes in this Figure correspond to categorical variables whereas the arcs represent the dependencies between these variables. In this figure only the parameters connected to MOE1 with arcs are used in the model. The model works by inferring the values of response variable using the joint conditional probabilities between MOE1 and eight predictor variables. For instance, given that all eight predictor variables are approximately uniformly distributed, as in Figure 21, the number of Red killed takes high values in 37.2 percent of the cases.

Bayesian Network Model of Red Killed

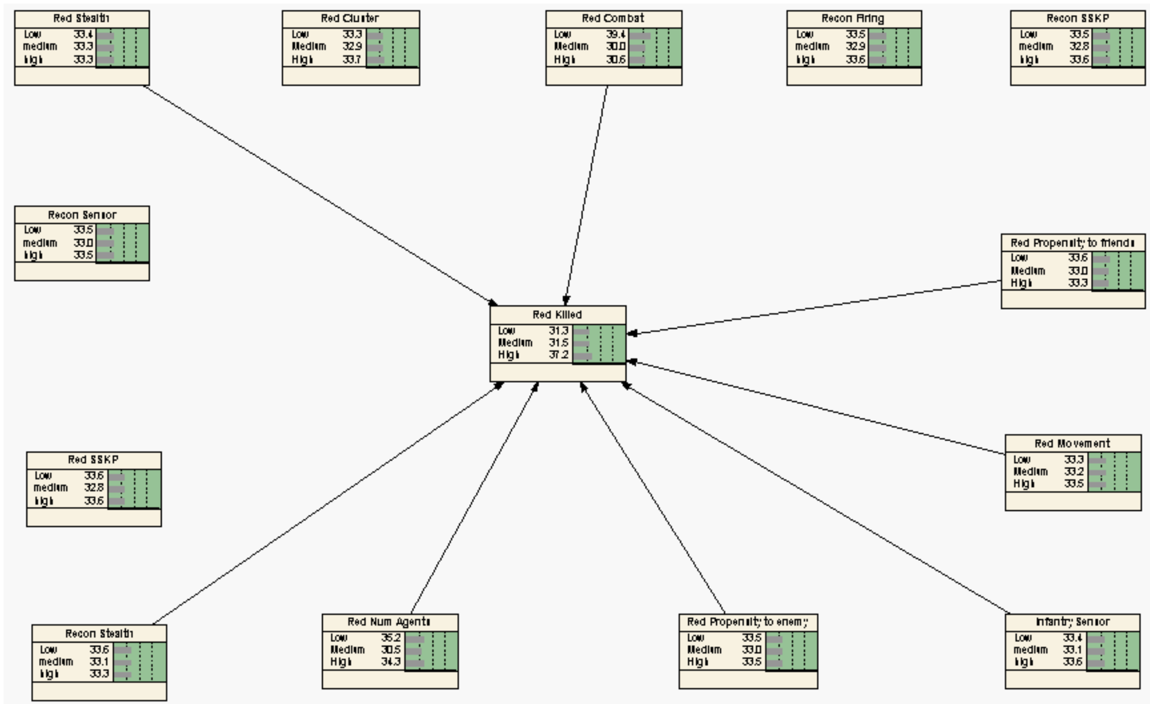

Figure 21. Bayesian Network Model of MOE1 for the 22-factor Data Set

The predictions are made using this model on the test set. The misclassification error rate is 0.026. Table 3 illustrates the predictions along with the actual values, the error rate and the variables used in the model.

.......**Predicted**......			
Low	Medium	High	Actual
619	28	0	Low
20	848	152	Medium
23	67	9545	High

Error rate = **2.566%**

Variables used in Model Construction

Red Stealth	Red Propensity to enemy
Red Combat	Infantry Sensor
Recon Stealth	Red Movement
Red Num Agents	Red Propensity to friends

Table 3. Bayesian Network Predictions of MOE1 for the 22-factor Data Set

b. Blue Killed

For MOE2 the same approach as MOE1 is used to build the model. The model in Figure 22 shows the network with nodes and arcs between nodes representing the relationships. Here in this figure given that all eight predictor variables are approximately uniformly distributed Blue takes low casualties 38.7 percent of the time.

Bayesian Network Model for Blue Killed

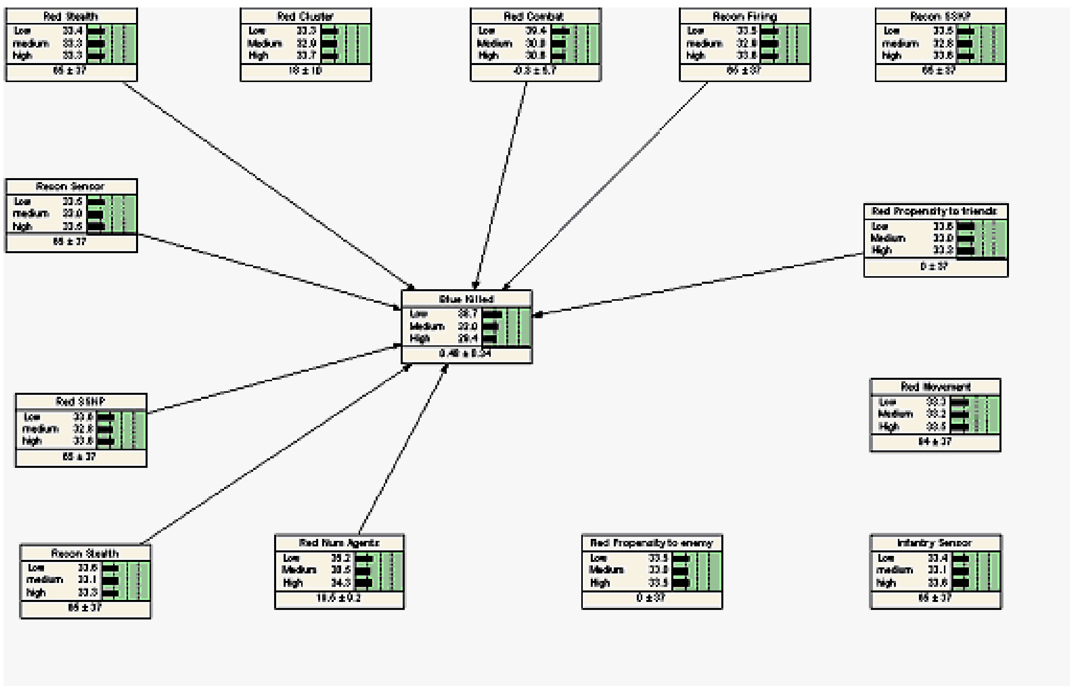

Figure 22. Bayesian Network Model of MOE2 for the 22-factor Data Set

After constructing the model, it is applied to the test set to see its predictive power. The misclassification error rate is 0.056, which is the model misclassified MOE2 in 5.6% of the test data. Table 4 shows the predictions against actual values and the variables used in model construction. In the table, the variables with an asterisk are the ones that also show up in the Bayesian Network model of MOE1.

.......**Predicted**......			
Low	Medium	High	**Actual**
------	------	------	------
8179	210	10	Low
325	2423	19	Medium
5	65	66	High

Error rate = **5.61%**

Variables used in Model Construction

*Red Stealth	*Recon Stealth
Recon Firing	Red SSKP
*Red Propensity to friends	Recon Sensor
*Red Num Agents	*Red Combat

Table 4. Bayesian Network Predictions of MOE2 for the 22-factor Data Set

2. The 6-factor Data Set

Since six predictor variables exist in this data set all of them are used in the model. The data set is divided into training and test sets. The training set is used to build the model and the test set is used to test it.

a. Red Killed

The model for MOE1 is shown in Figure 23. Apparently this model is fairly accurate in predicting the response with no error. In other words, given the values of these six variables and their relationships with MOE1, this model can predict the values of the response with perfect accuracy. As previously stated the model uses joint conditional probabilities between response and predictor variables to infer the values of the response variable. This model suggests that all of the six variables are important in predicting MOE1, as expected.

Bayesian Network Model for Red Killed

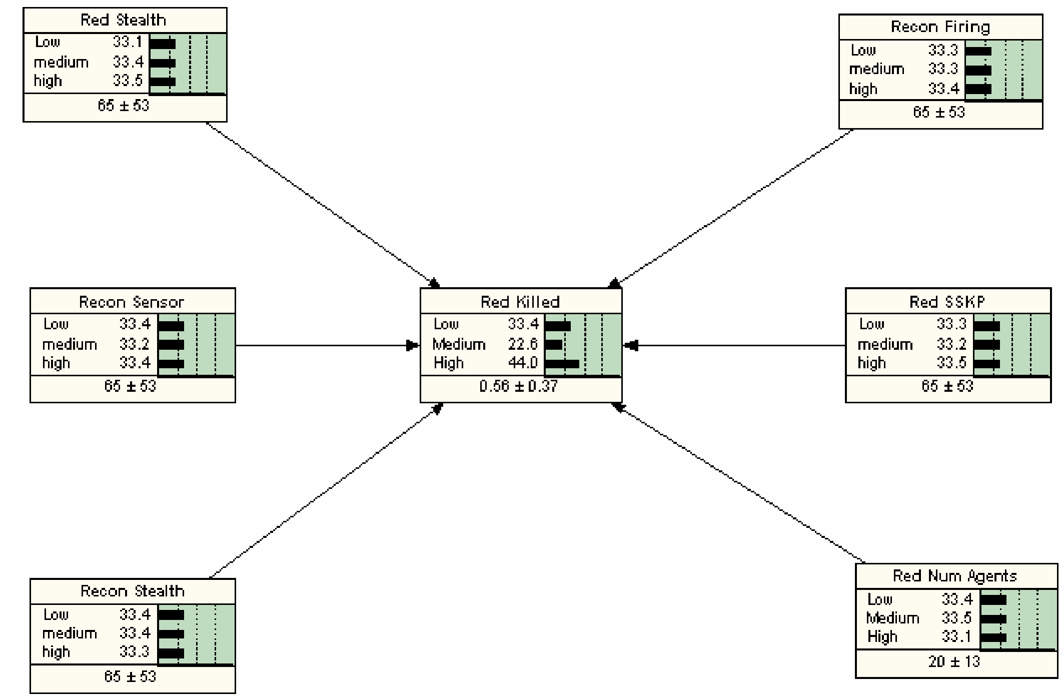

Figure 23. Bayesian Network Model of MOE1 for the 6-factor Data Set

Table 5 illustrates the predicted and actual values, and the misclassification error rate. As previously mentioned in this data set, the MOE1 takes either low or high values depending on different parameter settings. The Bayesian Network model seems to work even more efficiently in such cases where medium values of response variables almost never show up.

.......Predicted......			
Low	Medium	High	Actual
1398	0	0	Low
0	0	0	Medium
0	0	2902	High

Error rate = **0%**

Table 5. Bayesian Network Predictions of MOE1 for the 6-factor Data Set

b. Blue Killed

The model for MOE2 is shown below in Figure 24. The model learns the data using the training set and then uses probabilistic inference based on joint conditional probabilities to predict the values of response variable.

Bayesian Network Model for Blue Killed

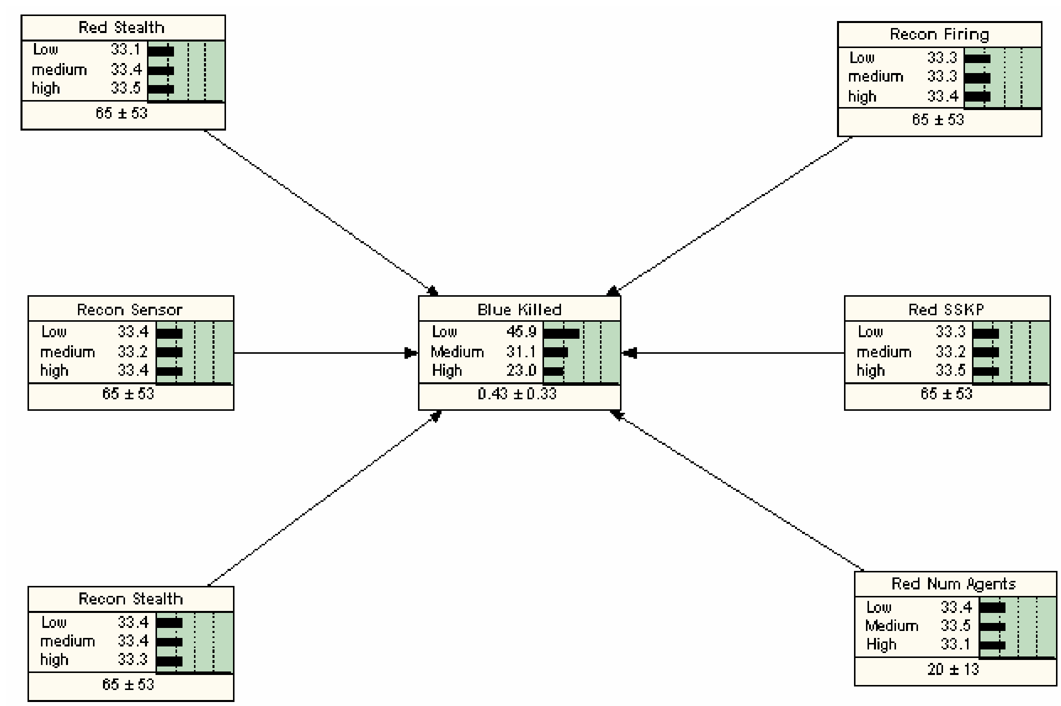

Figure 24. Bayesian Network Model of MOE2 for the 6-factor Data Set

In Table 6 the model for the 6-factor data set does much better than the one for the 22-factor data set in predicting the response variable with only 0.027 misclassification error rate.

.......Predicted......

Low	Medium	High	Actual
3152	18	0	Low
82	991	13	Medium
0	3	41	High

Error rate = **2.698%**

Table 6. Bayesian Network Predictions of MOE2 for the 6-factor Data Set

G. REGRESSION MODELS FOR THE 6-FACTOR DATA SET

Regression is the standard technique for assessing how various predictors relate to a response. This method is used to construct models for the 6-factor data set. Before building the model the data set is divided into training and test sets.

1. The Linear Regression Model for the Proportion of Blue Killed

Linear regression is used to describe the effect of continuous or categorical variables upon a continuous response. The linear regression model assumes that the response is obtained by taking a specific linear combination of the predictors and adding random variation (error) [Ref.29]. For MOE2, many transformations are considered but the best transformation is finally discovered to be the square root of the proportion of Blue killed. After this transformation, the StepAIC function is used to find the best linear model with main effects and two-term interactions. The StepAIC function is a special search method to find the most important main effects and interaction terms. It is usable in S-Plus by attaching library (Mass). Table 7 shows us the main effects and interactions in this model along with their coefficient values, standard errors, t-values and p-values associated with these t-values. Note that 12 two-term interactions appear in this model. For example, to predict the number of Blue killed in a certain situation, we take the value of main effects and interactions from the design and multiply them by their coefficients shown under the Value column. These values are added up taking the square of the summation. The signs of the coefficients of the factors are also worth looking at. For instance, the value of the Red Stealth's coefficient is (+0.0018) meaning that the number of Blue casualties increases as the Red Stealth increases. The residual standard error in this table illustrates how variable the errors of the predictions created by this model are. Multiple R-Squared is a measure of how good the regression model is in explaining the data. It shows how much of the total variation in the data can be explained by the regression model. The F-statistic compares the model to the model that has only the intercept. The null hypothesis is the terms except for the intercept do not help in predicting the response variables. If the p-value is less than our level of significance, we reject the null hypothesis. In order to see how powerful this model is when predicting

future data the test set tried. The residual mean deviance is 0.0296 on the test set. This value is simply calculated by summing the squares of the differences between actual and predicted values and dividing this by the number of observations.

a. Model

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0589	0.0069	-8.5832	0.0000
Red.Num	0.0008	0.0002	3.8700	0.0001
Red.Stealth	0.0018	0.0001	27.5267	0.0000
Red.SSKP	-0.0006	0.0001	-8.0395	0.0000
Recon.Sensor	-0.0007	0.0001	-10.6925	0.0000
Recon.Firing	-0.0007	0.0001	-10.8866	0.0000
Red.Stealth:Red.SSKP	0.0000	0.0000	83.6031	0.0000
Recon.Stealth:Recon.Firing	0.0000	0.0000	37.7593	0.0000
Recon.Stealth:Recon.Sensor	0.0000	0.0000	36.4957	0.0000
Recon.Sensor:Recon.Firing	0.0000	0.0000	27.6237	0.0000
Red.Num:Red.SSKP	0.0000	0.0000	21.4914	0.0000
Recon.Stealth:Red.Stealth	0.0000	0.0000	-15.7155	0.0000
Red.Stealth:Recon.Firing	0.0000	0.0000	-8.7754	0.0000
Red.Stealth:Recon.Sensor	0.0000	0.0000	-8.5192	0.0000
Recon.Stealth:Red.SSKP	0.0000	0.0000	-10.4685	0.0000
Red.Num:Recon.Stealth	0.0000	0.0000	-7.7536	0.0000
Red.SSKP:Recon.Firing	0.0000	0.0000	-5.1305	0.0000
Red.SSKP:Recon.Sensor	0.0000	0.0000	-4.8794	0.0000

```
Residual standard error: 0.1878 on 19982 degrees of freedom
Multiple R-Squared: 0.6757
F-statistic: 2449 on 17 and 19982 degrees of freedom, the p-value is 0
```

Table 7. Linear Model of MOE2 for the 6-factor Data Set

b. Normality Assumptions

In order for linear regression models to be valid, the error is assumed to have a Gaussian (normal) distribution with constant variance and to be independent of the predictor values. Noticeably in Figure 25 the residuals of our linear model have a constant variance while being distributed evenly around zero. However, this plot also shows that the model is still under fit (underdeveloped) and needs improvement. The plots in Figure 26 illustrates that the errors are approximately normal and symmetrical. In conclusion, the linear model above is a plausible model with normal and constant residuals.

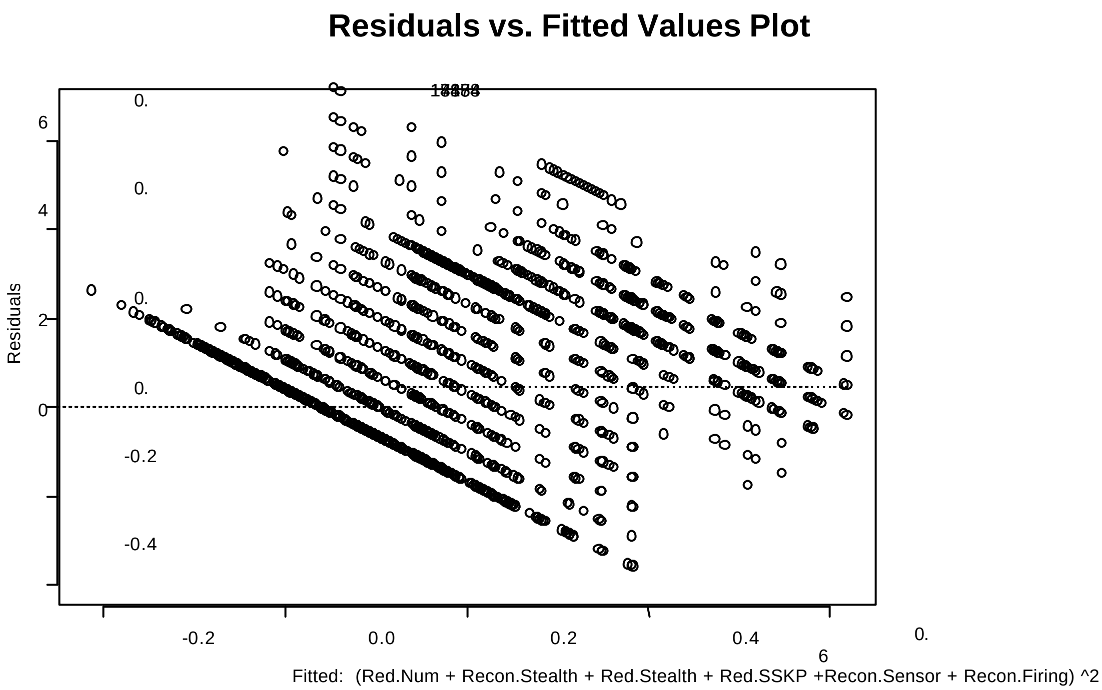

Figure 25. Residuals vs. Fitted Values for Linear Model of MOE2

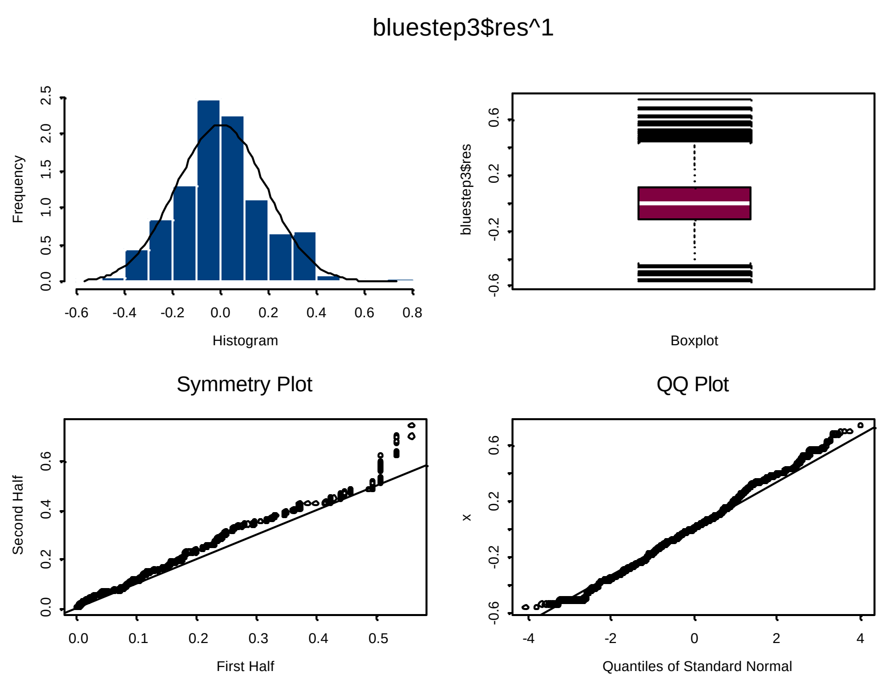

Figure 26. Normal and Symmetry Plots of the Residuals for Linear Model

2. The Regression Tree Model for the Proportion of Red Killed

After looking at the histograms and clusters associated with MOE1, we observed that this MOE take either low values, less than 0.25, or high values, greater than 0.75. Upon this discovery, a Regression Tree model seems appropriate for predicting the proportion of Red killed. This model reveals a very interesting and important fact that the values of MOE1 can be predicted with great precision by using only two of the predictor variables. Table 8 shows which variables are used in the model as well as the size and the residual mean deviance of the model. When applied to the test set, it gives us a residual mean deviance of 0.0035, which is almost the same as the training error value. Finally, the tree in Figure 27 is our model. This model strongly suggests that the values of the response variable can be estimated by using two variables. The model is also very simple and strong in predicting future data.

```
Variables actually used in tree construction:
"Red.Stealth"   "Recon.Stealth"
Number of terminal nodes:  3
Residual mean deviance:  0.003473 = 69.45 / 20000
```

Table 8. Regression Tree Model of MOE1 for the 6-factor Data Set

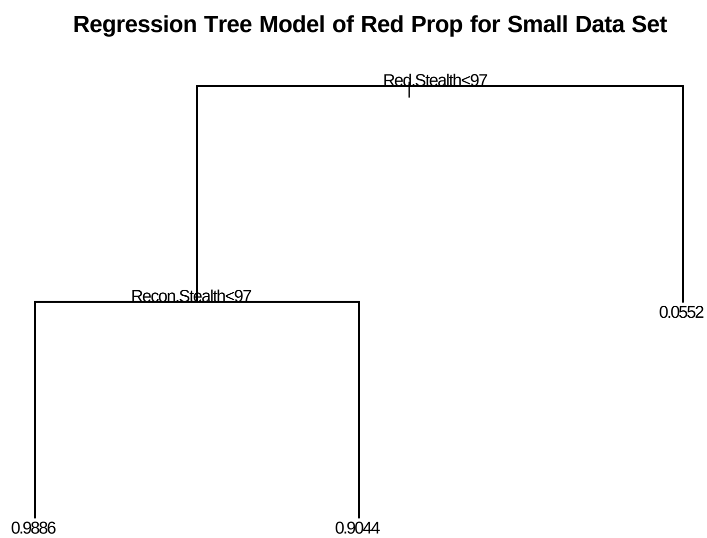

Figure 27. Regression Tree Model of MOE1 for the 6-factor Data Set

H. RESEARCH QUESTIONS

Prior to the analysis, seven research questions of interest were presented. In this section these questions are referred to by exploring the data sets using visual analysis tools and statistical methods. These techniques include Sign Tests, Trellis plots, Scatter matrices, and Bayesian networks.

1. The Blue Formation

The First sets of sign tests are performed to determine the difference between two different blue formations. For MOE1, the sign test shows no significant difference between the two scenarios. In other words, the different formation of Blue vehicles and infantry does not have a significant effect on the number of Red casualties.

For MOE2, the sign test suggests a significant difference between these two formations, and Scenario two, which has less Blue casualties, is superior to Scenario one. In other words, positioning vehicles up to the North and infantry down to the South is a better tactic for the Blue side in this specific infiltration scenario in terms of Blue casualties.

2. The Red Formation

For this question, the focus is on the third infiltration team of the Red side. Basically, we want to know whether a third infiltration team would make a difference on the MOEs. Looking at the results of the sign test for MOE2, this three-team formation of the Red does not have a significant effect on Blue casualties. On the other hand, this configuration does cause the Red side more casualties, increasing their chance of losing agents. In conclusion, in this specific scenario, the three-team formation is not a better tactic than the two-team formation for the Red side because by using the three-team formation, the Red side takes more casualties than the Blue side.

3. The Number of Red Agents

The third question addresses the optimal number of Red agents to negate the Blue superiority in number and the ability of the agents. To answer this question, Trellis plots

and Bar charts are used. Figure 28 is the plot of MOE1 vs. MOE2 conditioned on the values of the number of Red agents. In this plot the goal is to find the best condition for Red that is low value for the Red Prop and high value for the Blue Prop. When looking at the plot on the upper left corner, we see a cluster of points, which are in favor of Red. Figure 29 also gives an idea of the best size of the Red teams. This is a 3-D Barchart of the Blue Prop against the Red Stealth and the Red NumAgents. This plot suggests that the optimal size of the Red teams should be around 16 to 24. From these observations it can be concluded that the best number of Red agents for each Red infiltration team should be around 16 to 27.

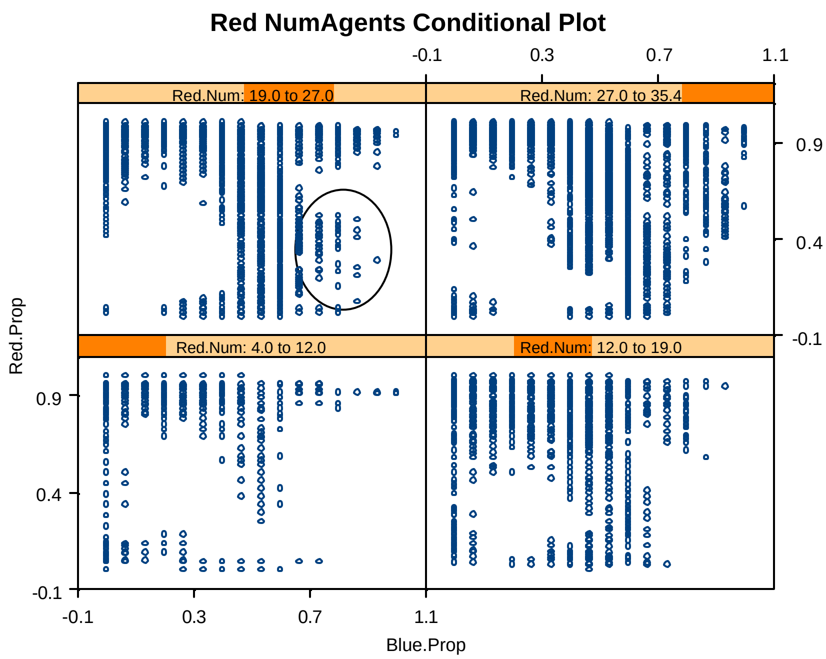

Figure 28. Trellis Plot of Red Number of Agents Parameter

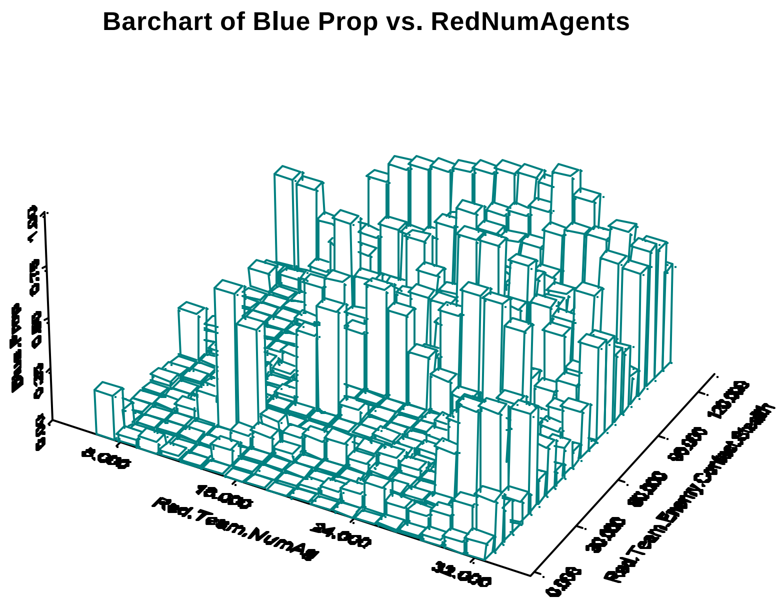

Figure 29. Barchart of MOE2 vs. Red NumAgents and Red Stealth

4. Aggressiveness of Red Agents

This question relates to the Red personality weightings, w2, w8, w10 and the Red combat constraint. The focus is on the values of these parameters and their effects on our MOEs. To be able to answer this question Netica Bayesian Networks and Trellis Plots are used. Figure 30 illustrates that if the Red agents remain indifferent against Blue vehicles and commanders, the Red agents do not attack the enemy and escape from the Blue infantry, the Blue side takes the minimum number of casualties.

Bayesian Network of Red Personalities (Worst Case)

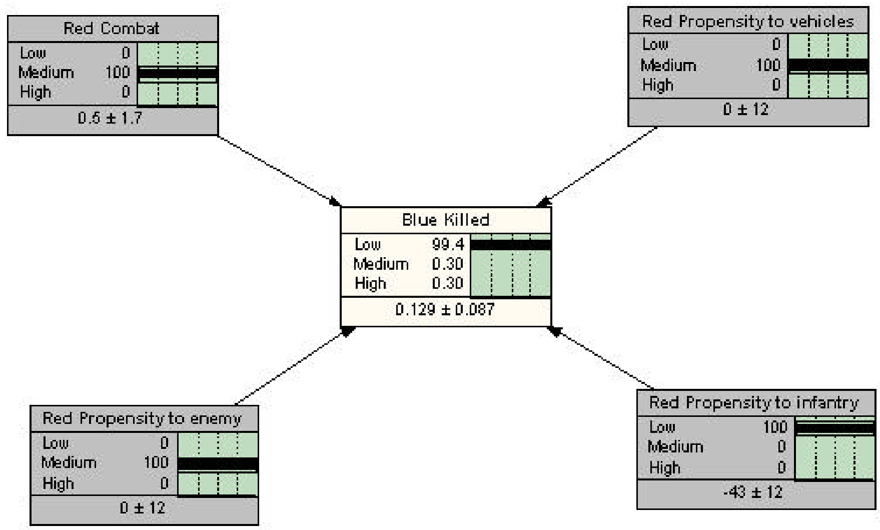

Figure 30. Bayesian Network of MOE2 for Red Personality Changes (Worst Case)

In contrast, Figure 31 suggests that if Red is aggressive, Red's propensity to the Blue commanders and vehicles is high and Red's propensity to the Blue infantry is low, Blue takes medium and high casualties more often.

Bayesian Network of Red Personalities (Best Case)

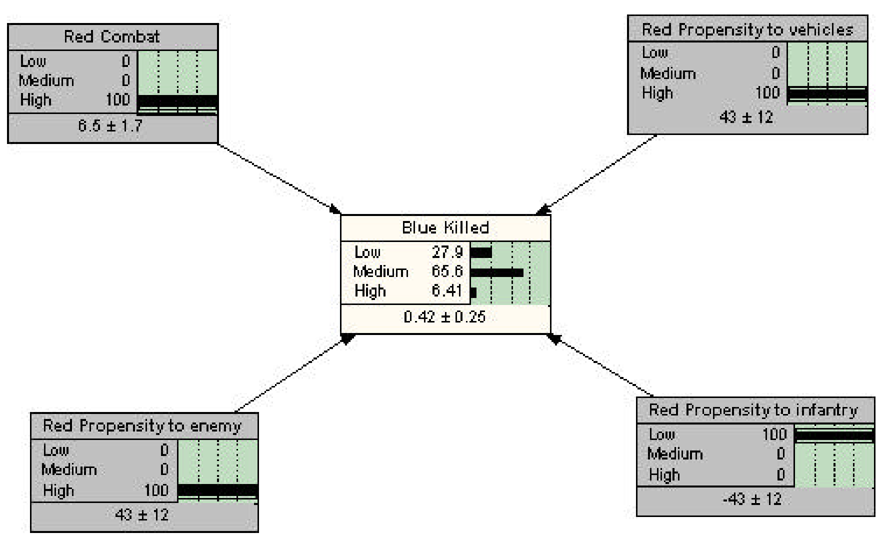

Figure 31. Bayesian Network of MOE2 for Red Personality Changes (Best Case)

Figure 32 shows the best case for Red's personality in terms of Red casualties. If the Red side is unconcerned about the Blue infantry and vehicles, and is not aggressive but fearful about the Blue commanders then the Red side does better by taking less casualties than other times.

Bayesian Network of Red Personalities (Best Case)

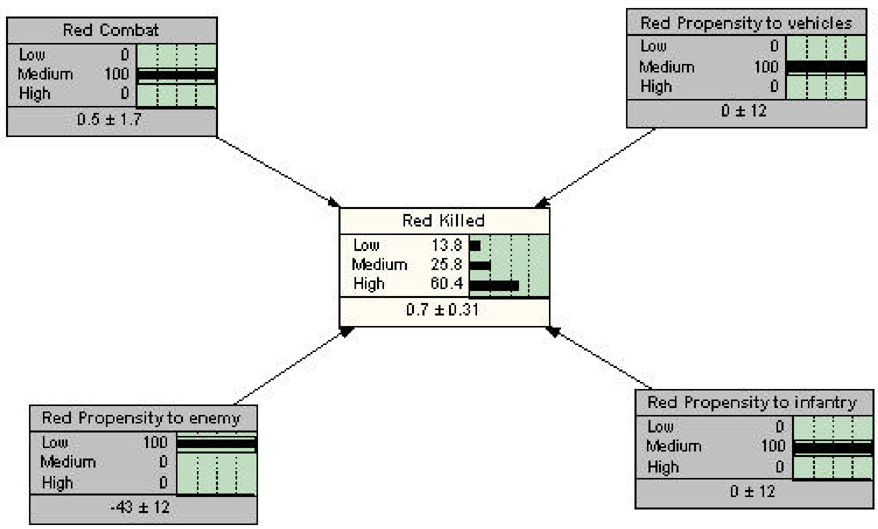

Figure 32. Bayesian Network of MOE1 for Red Personality Changes (Best Case)

Figure 33 illustrates the worst case for the Red team. If the Red agents remain indifferent against Blue vehicles and commanders, do not attack the enemy and run away from the Blue infantry then Red takes the maximum number of casualties.

Bayesian Network of Red Personalities (Worst Case)

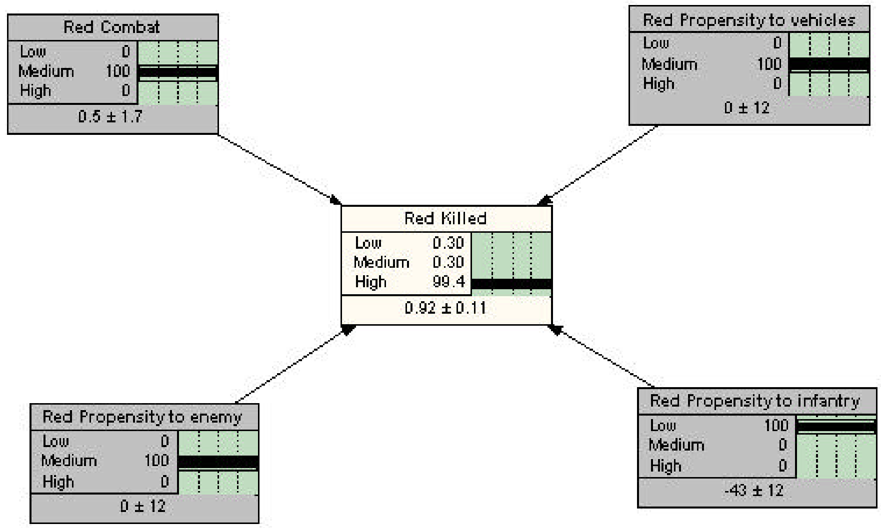

Figure 33. Bayesian Network of MOE1 for Red Personality Changes (Worst Case)

Trellis plots (Figure 34) shown below support the findings from Bayesian Networks. The first plot for the Red Combat shows that if Red agents become more aggressive then the Red side takes low or medium casualties while more often inflicting high Blue losses. This is evident by especially looking at the points inside the circle. In Figure 35 also the plot on the upper right corner suggests the best results for the Red team. In this plot, the best results are obtained when the Red propensity to the Blue vehicles and infantry is high.

In conclusion, the Red team gets the best results in terms of the proportion of Blue killed when being aggressive, attacking the Blue vehicles and commanders, and repelling from the Blue infantry. If MOE1 is the concern then Red does better when repelling from the Blue commanders while remaining indifferent to the vehicles and infantries. On the other hand, the worst result for the Red team obviously occurs when they are running away from the Blue infantry and when they are unconcerned about all other enemy agents.

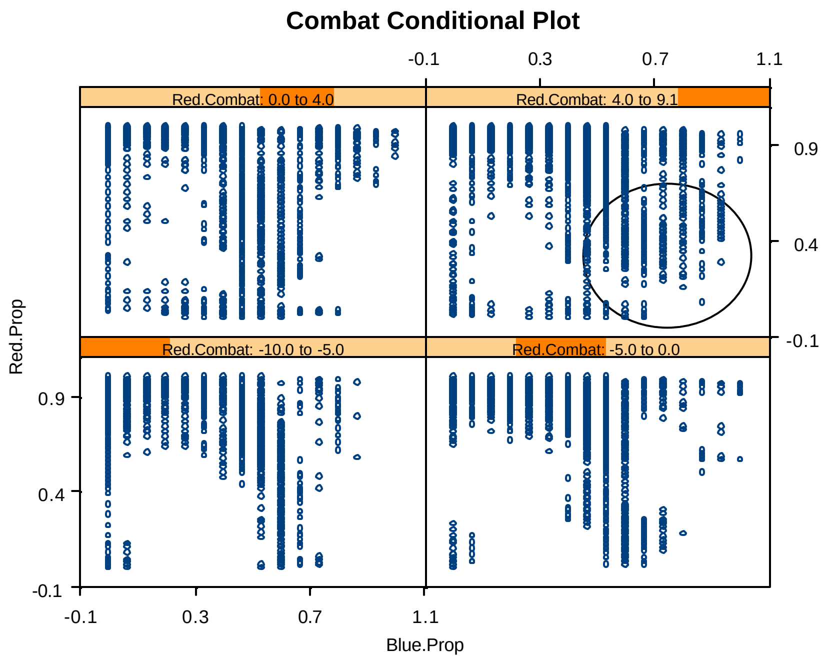

Figure 34. Trellis Plot of Combat Parameter

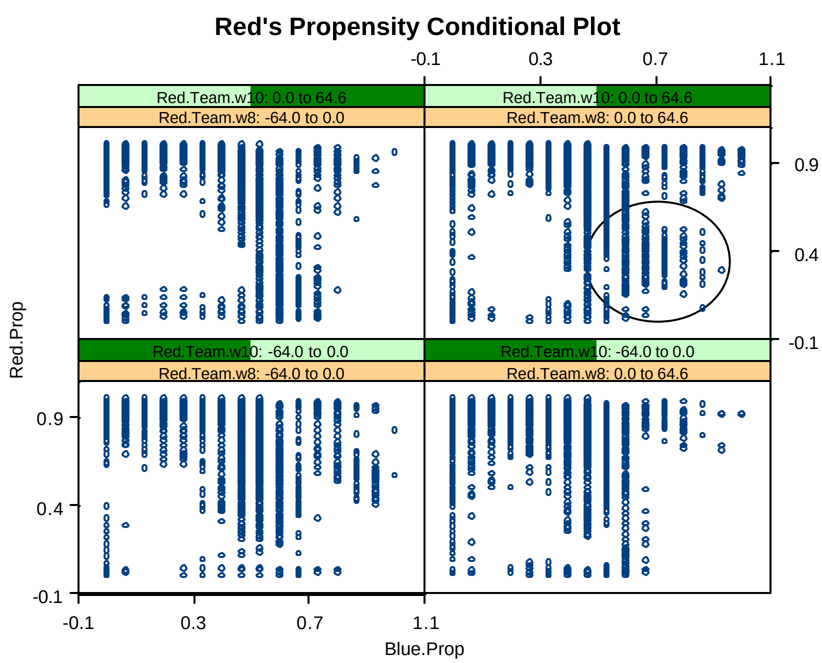

Figure 35. Trellis Plot of Red’s Propensity to Blue Parameters

5. The Optimum Number of Targets for Reconnaissance Team

One of our research questions is the optimum number of targets that should be fired upon by the Recon Team to help the infiltration. The network in Figure 36 says that the best results can be obtained for the Red side by this combination of Recon parameters. The Red force causes more casualties to the Blue force when the Recon Sensor and Firing Ranges are high, the Recon SSKP is high (shoots better), and the number of targets is medium. Medium values vary between 6 and 10 for the Recon Max Targets parameter.

Figure 37 also gives a sample of the optimum number of targets assigned to the Recon team. When the Recon Sensor range is high the values of Max Targets parameter vary between 8 and 15. If we combine these two results, we can conclude that the optimum number of targets to be assigned to the Recon team is around 8 to 10.

Bayesian Network of Recon Parameters (Best Case)

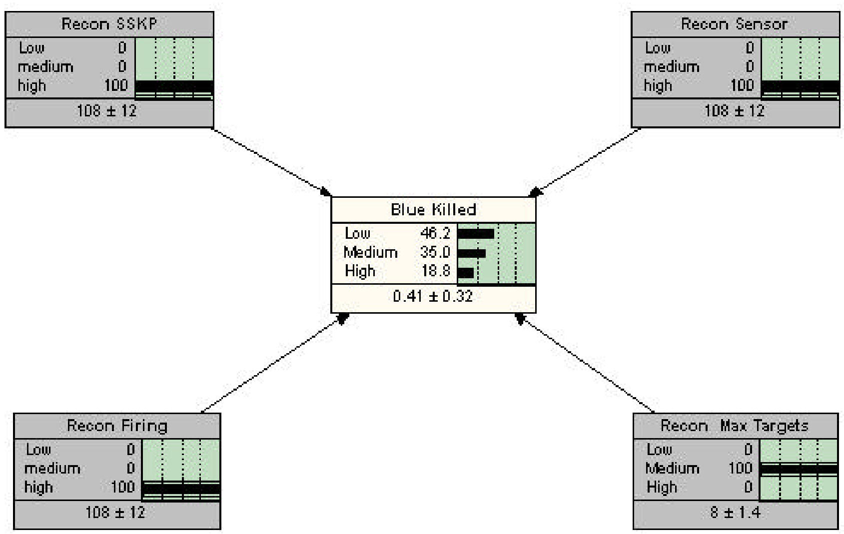

Figure 36. Bayesian Network of MOE2 for Recon Parameters (Best Case)

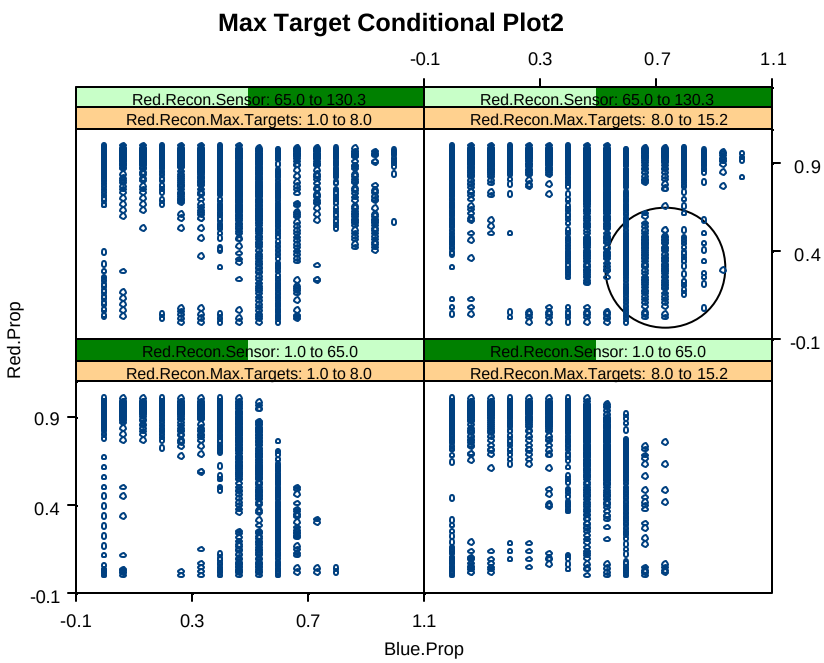

Figure 37. Trellis Plot of Recon Max Targets Parameter

6. Blue Infantry Parameters

Our sixth question relates to the Blue infantry parameters and, specifically, to the Stealth parameter. Belief networks and Trellis plots are used to explore the data to find an answer to this question. The network in Figure 38 implies that the best result in terms of Red casualties happens when the Infantry Stealth is high and all other Infantry parameters are low. This result is interesting because intuitively one expects it to be just the opposite. The data, however, suggests that if the Blue infantry remains stealthy, shoots poorly, and keeps their ranges low, the enemy loses more agents. This can be the result of Red agents getting closer and being killed by the Blue vehicles. In other words, since the Red agents get closer to engage with the enemy they become the targets of the Blue vehicles. Moreover, since the Red agents engage longer with the Blue agents due to poor shooting and high stealth of Blue infantry they suffer more casualties. The same result can be seen in Figure 39. In the lower right plot in this figure, the best results occur for the Blue side where the Red side continuously takes high casualties.

Bayesian Network of Infantry Parameters (Best Case)

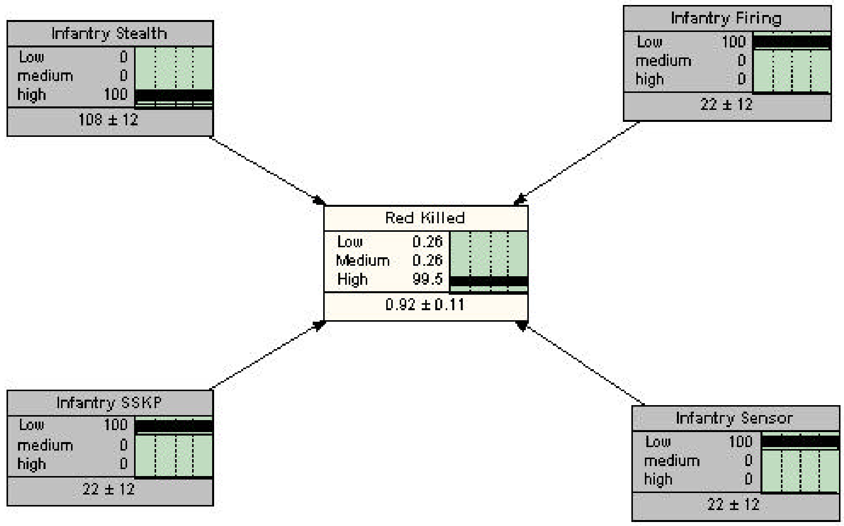

Figure 38. Bayesian Network of MOE1 for Infantry Parameters (Best Case)

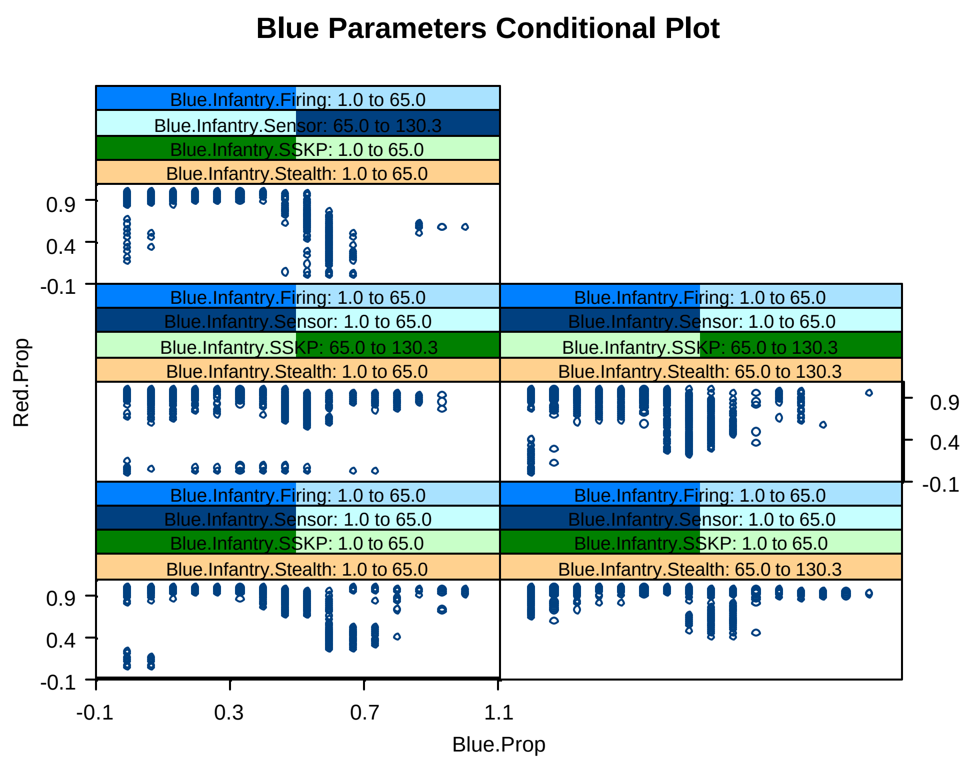

Figure 39. Trellis Plot of Blue Infantry Parameters

7. The Red Cohesion

Our final question is the effect of Red's cohesion on the outcome of the battle. To measure this effect, three parameters of the Red force, the Red w1 (Propensity to friends), the Red w3 (Propensity to injured friends), and the Red Cluster, are used. The first network in Figure 40 suggests that the Red side causes more casualties to the Blue side when the Red agents are not attracted to their friends and when they are allowed to form big clusters.

Figure 41 shows the best outcome for the Red team in terms of MOE1. Here the Red team takes fewer casualties if the Red agents repel from their injured friends, remain indifferent to their friends, and form big groups to attack the enemy. Figure 42 also gives us similar results. In this Figure, in the plot on the upper right, we notice the group of points that favor the Red side. This plot says that the best tactic for the Red team is to form big clusters, go after friends and repel from injured friends. In conclusion, more cohesive Red agents attracted to their friends and repelled from injured friends always perform better in this specific scenario.

Bayesian Network of Red Cohesion Parameters for MOE2 (Best Case)

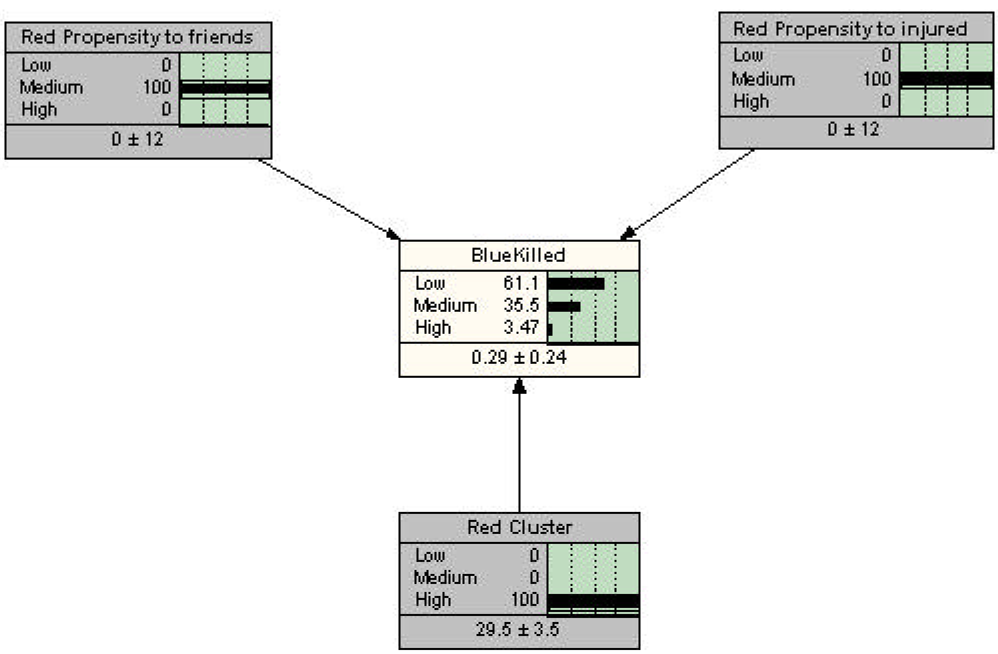

Figure 40. Bayesian Network of MOE2 for Red Cohesion Parameters (Best Case)

Bayesian Network of Red Cohesion Parameters for MOE1 (Best Case)

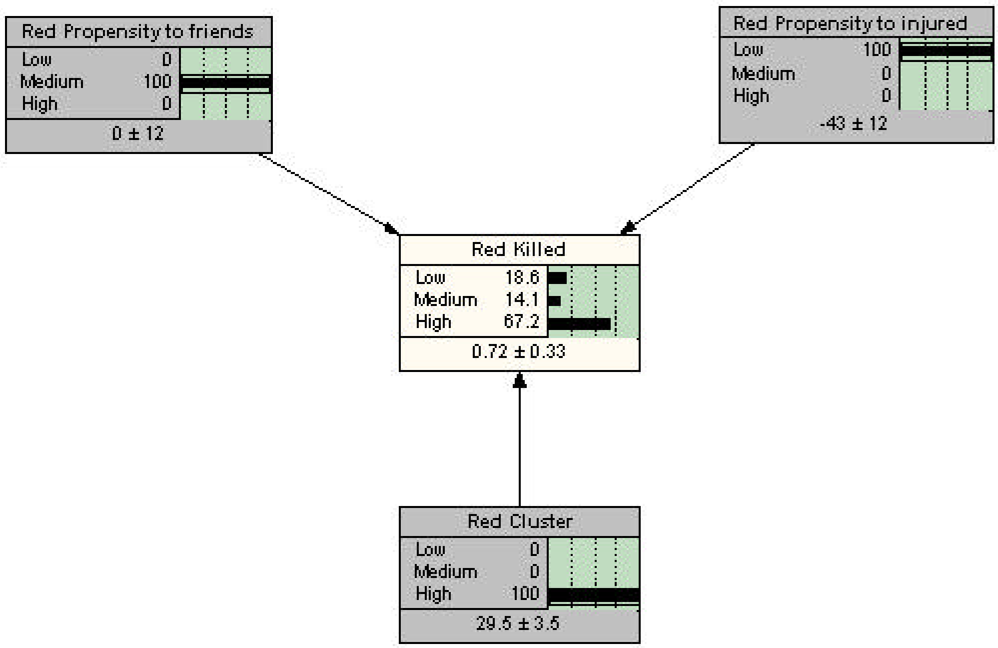

Figure 41. Bayesian Network of MOE1 for Red Cohesion Parameters (Best Case)

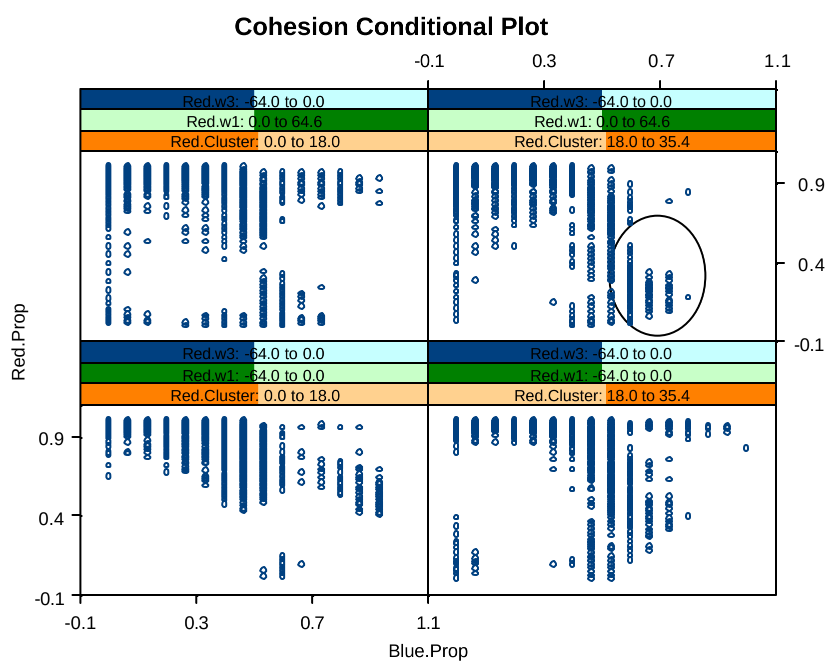

Figure 42. Trellis Plot of Red Cohesion Parameters

In this chapter, the data sets are explored using several statistical analysis techniques and data visualization tools. Several models developed using various statistical methods are presented. The focus is first on describing the data and then using it to predict the future data. The most important parameters for predicting MOEs are discovered and mentioned. Finally, this chapter is concluded by answering our research questions based on the visual exploration of the data sets. In the next chapter, we will give some recommendations to the model developers of MANA and the analysts who will study further in this field.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. RECOMMENDATIONS

"So there are five ways of knowing who will win. Those who know when to fight and when not to fight are victorious. Those who discern when to use many or few troops are victorious. Those whose upper and lower ranks have the same desire are victorious."

Sun-Tzu

A. CHAPTER OVERVIEW

This chapter explains the recommendations to the developers of MANA and to the analysts that would like to study in this field. The focus is on the possible improvements that might be considered by the model developers and the difficulties that have been encountered throughout the process of this thesis.

B. MODEL DEVELOPERS

This section explains the problems that have been encountered in the phase of putting the scenarios in MANA. It also discusses some issues for consideration in the next versions of the model.

- Terrain features should be considered in more detail. In real situations, especially in guerrilla warfare, terrain is one of the most important factors that affect the outcome of the battle. Because of this, terrain features, such as woods, buildings, hills, visibility, line of sight, should be incorporated into the model.

- Scenario files that can be imported across different versions of MANA should be considered as an improvement to future versions. Currently all available versions of MANA do not allow this.

- Source code should be improved to overcome the problem of agents getting stuck due to obstacles on the battlefield. The agent's move selection algorithm could be modified to avoid this problem.

- In real situations agents of the same squad must be placed in the same location when taking up a defensive position on the battlefield. They should stay put and stick to other agents in their squad. In MANA since agents are not allowed to occupy the same cell, one must insert the agents individually in the model. Otherwise, the agents in the same squad wander around their position and never stand still.
- Vehicles, like tanks or ACVs, are not exactly realistic in the model. A great deal of effort is needed to achieve the real performance of these agents. Moreover, it is almost impossible to consider the abilities of the vehicles' crewmembers. Better algorithms should be incorporated to distinguish vehicles from other agents. For example, a tank has three weapons with completely different capabilities and four crewmembers with different tasks. Thus, it should be modeled as a more complicated system than regular agents.
- Lethality of the agents can also be incorporated in MANA as a personality parameter.
- Guerrilla units are eventually affected by strain, fatigue, and stress in guerrilla-style operations. Integrating some parameters associated with these intangibles would be very useful to evaluate their effects on the outcome of an asymmetric war.
- Guerrilla battles that were looked at in this study emphasize the significance of night combat and combat under poor weather conditions. In most cases the guerrillas prefer to engage with conventional forces under these conditions. Therefore, weather conditions should also be considered in the model.
- Mine locating and clearing is one of the most crucial concerns, awaiting the units encountering low intensity conflicts. Hence, mine warfare should be considered as one of the future enhancements to the model.

- An increase of MOEs should be implemented including time to mission completion, number of enemy agents seen, etc.

- Near Orthogonal Latin Hypercube Designs developed by Tom Cioppa allows for a structured way of exploring MANA. Integrating these designs in Supercomputer would be useful. Furthermore, fractional factorial designs can also be incorporated to study the effects of the variables and their interactions.

- Bayesian Belief Networks provide a means of presenting the effects of several variables simultaneously. They offer an insightful way of displaying the effects and interactions of many variables in an organized and understandable format. Incorporating Bayesian Networks in further developments of Agent Based Models would be beneficial.

C. ANALYSTS

This section gives some recommendations to the analysts who would like to study in this field in the future.

- Great care should be given for developing a scenario and incorporating it in the model. Insuring that the scenario runs on the model similarly to how it would in the real-life is critical.

- Studying the effects of all variables in an ABM is almost impossible in most cases. Thus, a clear understanding of which variables are of interest and how they will vary is necessary.

- After the variable selection process the design to be used to explore the data should be determined and developed.

- Several analysis techniques can be used to explore the data. From our experiences, Cluster Analysis, Neural Networks, Regression Trees, and Bayesian Belief Networks are good tools when so many variables are of interest. Also Trellis Plots and Scatter Matrices provide an insightful visualization of the data for the effects of various factors.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. SUMMARY OF THE BATTLES

A. OVERVIEW

This Appendix introduces six guerrilla battles, which were fought after the WWII. The Appendix includes a brief summary and the lessons learned in terms of guerrilla tactics and techniques for each battle.

B. THE CHINDITS

The decisive campaign against the Japanese in Burma in World War II was the Battle of Imphal-Kohima. For 100 days, from March through June 1944, the troops of the Fourteenth Army met their Japanese enemies in an intense struggle for control of the eastern gates to India. Ultimately, this British Army defeated the Japanese and began the slow task of clearing the invaders from northern and southern Burma. While this great battle was being contested, another war was being fought 200 miles in the Japanese rear. In this conflict, over 20,000 specially-trained jungle soldiers attempted to weaken the Japanese Army by delivering a “knockout blow to its unprotected guts”. Three-thousand of these troops were American volunteers, officially known as the 5307th Composite Unit or more popularly known as “Merrill’s Marauders” or “Galahad”. The other larger part of these soldiers was the Chindits, also known as the Special Force. Essentially, Galahad and the Chindits were light infantry jungle troops organized and trained for guerrilla-style interdiction against Japanese lines of communication (LOC) [Ref.8].

Toward the end of the war, taking a lesson from the Chindits, the Fourteenth Army went on to defeat the Japanese Army in Burma in the Chindit style of boldness, aggressiveness, and confidence [Ref.8].

By the end of 1943, the Japanese had given up on invading India, believing that the jungles beyond the Chindwin River in Burma were impassable. The Japanese choose to remain static and protect what they had. The British were also more or less content in defending India [Ref.38].

1. Lessons learned

The Chindit War has great value for military historians and analysts because of the many conclusions regarding guerrilla-style operations in an enemy's rear. These conclusions are:

- A properly-trained unit in the rear of an enemy can have an effect highly un-proportion to the actual numbers of men involved. Furthermore, a small force can defeat a larger force if it achieves surprise and attacks the enemy where it least expects an attack.
- Guerrilla forces would prove themselves superior to their enemies by training hard and adapting to the environment. They should modify their tactics to exploit the terrain better than their foes.
- Galahad operations have exhibited a number of tactics that remain a model for guerrilla warfare: "a swift approach march along an unguarded route"; the preservation of surprise; a hasty, accurate reconnaissance, followed by a bold attack against the enemy's weakness; and the employment of well-aimed, disciplined fires [Ref.8].
- Guerrilla units are eventually affected by strain, fatigue, and stress. A limit exists as to how long a unit can be expected to remain effective in the enemy's rear. Evacuation of such men after a certain amount of time seems to be a targeted action.

C. THE CHINESE COMMUNIST FORCES (CCF) IN KOREA

In the autumn of 1950, forces of the United Nations Command, directed by General Douglas MacArthur, pushed confidently through the mountains of North Korea toward the Yalu River and the Manchurian border. The UN Command advanced optimistically, even daring to forecast an end to the war by Christmas. They did not know that a huge Chinese Army lay in wait, tensing for the right moment to overcome the UN columns.

The Chinese Army's dependency on manpower and the roughness of the Korean terrain determined how the CCF was structured and employed during the Korean War. Compensating for its weakness in armaments and exploiting the possibilities of the rough Korean landscape, the Chinese developed a philosophy of "man over weapons" organizing a light infantry army to fight the war.

In the first year of the war, the Chinese frequently demonstrated superior field craft, an almost inhuman endurance, and a sharp appreciation for terrain. Undaunted by weather, terrain, or poverty, the CCF, during this stage of the war, pressed the UN Command to its limits.

The most obvious weakness of the CCF was its severe shortage of military equipment. Furthermore, the UN in firepower, transport, and air power hopelessly outmatched the Chinese. A further Chinese debility was their tactical rigidity. Ultimately, these weaknesses were all magnified during the last stages of the war and the Chinese advantages in tactical maneuver, infiltration, and stealth lost their values. By the autumn of 1951, the CCF lost its psychological advantage over the UN forces and began to suffer a morale problem. The firepower of the UN had equalized the manpower imbalance and, in the final analysis, negated Chinese strengths. In conclusion, the Korean War also represents the limits to which the "man over weapons" philosophy can be carried [Ref.8].

1. Lessons learned

The Korean War exhibited very important aspects of guerrilla warfare. These are the most important lessons that can be taken from this war:

- The main strengths of the CCF in the Korean War were its powerful philosophy of "man over weapons," the skills and abilities of the individual light infantrymen, and the effectiveness of the CCF leadership. The integration of these strengths created a fierce battlefield instrument that achieved remarkable tactical successes [Ref.8].

- Another lesson that can be drawn is that improvisation is a very crucial issue for exploiting asymmetric wars. Guerrilla forces should use

whatever resources are at hand for the military purposes of camouflage, deception, booby traps, fortifications, and sustenance.

D. BRITISH OPERATIONS IN MALAYA AND BORNEO

1. The Malayan Emergency

From 1948 to 1966, substantial British, Gurkha, and Commonwealth infantry forces participated continuously in prolonged light infantry operations in the Far East. In Malaya, from 1948 to 1960, these British-directed forces defeated an indigenous Communist insurgent force. Less than three years later, the British Army moved into North Borneo to secure that territory against Communist guerrillas and Indonesian aggression in a four-year war. In both wars, the combat took place in extremely inhospitable terrain and it was swift, fleeting, and violent.

In Malaya, the Communist insurgency had its origins in the organizations established by the Malayan Communist Party (MCP) during World War II to fight the Japanese. Trained, armed, and supplied by the British, the military arm of the MCP-known as Malayan People's Anti-Japanese Army (MPAJA) grew into an extensive and efficient organization, some elements of which operated under British liaison officers. After the war ended, the MPAJA was disbanded and supposedly disarmed. However, the hard-core Communist elements of this small army hid their arms and supplies in secret caches for future use.

For a time, the MCP cooperated with the reestablished colonial administration in Malaya. When it became clear, however, that the aims of the MCP to influence the establishment of a socialist-type "People's Government" had no chance for success, the MCP adopted a more violent policy of social de-stabilization through labor unrest, strikes, and eventually, armed uprisings and acts of terrorism. In response, the government of the Federation of Malaya declared a state of emergency on June 18, 1948 and adopted emergency powers to deal with the violence. In addition, the MCP was outlawed on July 23, 1948 [Ref.8].

By 1960 the previously large terrorist jungle army was defeated and reduced to a few hundred men near the Thai border. The Malay Government declared the end of the Emergency in July 1960. The result of the war was the establishment of the independent democratic nation of Malaya, later to become the present day Malaysia [Ref.39].

2. The Confrontation with Indonesia

Not long after cleaning out the last isolated pocket of Communist guerrillas in Malaya, the British forces in the Far East found themselves facing another limited war, this time in North Borneo.

The large island of Borneo in 1962 comprised four political entities. Kalimantan, the southern three-fourths of the island, belonged to Indonesia, independent since 1949. In the North, Britain administered the two provinces of Sarawak (in the West) and Sabah, also called North Borneo (in the North). The sultanate of Brunei was an independent state ruled by a sultan but possessing a civil bureaucracy and police force staffed to a large degree by Englishmen.

In 1961, Tunku Abdul Rahman, the prime minister of Malaya, proposed the formation of a new federated state to be known as Malaysia. Malaysia, Rahman suggested, should include the Federation of Malaya, the city-state of Singapore, the sultanate of Brunei, and the colonial provinces of Sarawak and Sabah. Great Britain endorsed the idea, but President Sukarno of Indonesia opposed it, calling it a British neocolonialist project and a threat to Indonesian security. Sukarno had his own dreams about a greater East Asian Federation under his leadership, which also was to include northern Borneo. He openly announced a policy of "confrontation" in January 1963, following the Brunei revolt of December 1962. Thus, the Brunei revolt was actually the beginning of a series of conflicts known as the Confrontation. A small indigenous Communist organization, the Northern Borneo National Army (TNKU), which had ties to the Communist party of Indonesia, launched this revolt [Ref.8].

On the night of the September 30, 1965 a coup attempt, in which Indonesian communists became involved, occurred in Indonesia. Six senior generals were killed. The coup failed and was followed by widespread violence and bloodshed. It proved a

turning point for the Confrontation, which declined thereafter and on August 13, 1966 a formal agreement concluded between Indonesia and Malaysia bringing the conflict to an end [Ref.40].

3. Lessons learned

The British operations in Malaya and Borneo had their own unique sets of conditions and parameters that strongly influenced the flow of events as well as the methods and techniques that characterize the guerrilla operations in general:

- At the tactical level, the most important principle to be inferred from the British experience is that guerrillas must be masters of their environment. They have to endure tough terrain, exhaustion, heat and humidity, jungle pests, and severe mental and physical strain just to be able to get at the enemy. Additionally guerrillas have to be more adept and capable in the jungle than their enemies in order to dominate the environment whether it is jungle, mountain, or arctic territory. To achieve this mastery over the environment requires, above all else, a singleness of purpose, an attitude of self-reliance, persistent mental discipline and self-denial. Furthermore, such domination of the environment depends on the adoption of an offensive, aggressive policy [Ref.8].

- The development of timely, accurate intelligence leads to the success of tactical operations. This intelligence originates largely from the local inhabitants and the police organizations. Good intelligence appears to be a crucial cornerstone for guerrilla operations in counterinsurgency and low-intensity conflicts.

- The small-unit actions of the Emergency and Confrontation placed a premium on the quality of low-level infantry leaders. The burdens of responsibility and decision-making borne by the NCOs, lieutenants, and captains in these campaigns far exceeded that experienced by the same ranks in conventional large-unit operations.

- These conflicts emphasized once again that there should be a limit to the physical and mental strain placed on guerrilla fighters. After certain time the teams rapidly lost their effectiveness as their physical and mental abilities were worn out.

- The actions in Malaya and Borneo paralleled each other in important respects. For example, the British adhered to a 3x3 squad organization in both conflicts; therefore, they split their squads into 3 teams with 3 infantrymen in each team. This type of organization appeared to be highly valuable in independent squad-level actions, probably because of the flexibility that it provided [Ref.8].

- The development of good British marksmanship in both struggles was essential to success. Winning and surviving for a guerrilla meant shooting first, shooting straight, and shooting to kill.

- In both battles the best ideas on tactical operations come from the units themselves. The Director of Operations should put their staffs to work collecting and evaluating these ideas, not dreaming up their own notions and imposing them on the infantry units.

- The British approach to "winning hearts and minds" deserves a great deal of study among all military analysts and historians. This principle was vital to success at every level and was impressed on the lowest soldier in the chain of command. For infantrymen, it basically meant protecting the citizens of the country, respecting their persons and their property, and not being afraid to get close to them [Ref.8].

E. AFGHAN-SOVIET WAR

Soviet forces came to the assistance of a Marxist government that had come to power in Kabul through a military coup in April 1978. The hostility of the revolutionary government toward the Islamic religion led to increasing opposition. Soviet intervention on 24-27 December 1979 installed another Marxist leader in power. However, the Soviet

presence quickly led to the emergence of the mujahideen (the warriors of God), who waged a guerrilla war against Soviet and government forces. The Soviet Army found it hard to operate against Afghan guerrillas in the mountains, even when deploying modern airpower and helicopters against them. The mujahideen had a safe refuge in Pakistan and received modern weapons, including Stinger anti-aircraft missiles from the United States. International opposition and the rising cost of the war, including an estimated 13,000 dead, led to the Soviet's withdrawal in February 1989 and the fall of Kabul to the mujahideen in April 1992 [Ref.19].

1. Lessons learned

Modern, mechanized forces are still in peril when committed to fight guerrillas in the middle of a civil war on rugged terrain. The Soviet-Afghanistan war demonstrated the following Lessons-learned:

- A guerrilla war is not a war of technology versus peasantry. Rather, it is a contest of endurance and national will. The side with the greatest moral commitment (ideological, religious or patriotic) will hold the ground at the end of the conflict. Battlefield victory can be almost irrelevant, since victory is often determined by morale, obstinacy and survival [Ref.19].

- Secure logistics and secure lines of communication are essential for the guerrilla and non-guerrilla force. Security missions, however, can tie up most of a conventional force. Existence of safe havens, such as neighbor countries, is also a great asset for guerrilla forces to restructure and re-supply their units

- Weapons systems, field gear, communications equipment and transport that are designed for conventional war will often work less effectively or fail completely on rugged terrain.

- Tactics for conventional war will not work against guerrillas. Forces need to be reequipped, restructured and retrained for fighting guerrillas or for fighting as guerrillas. The most effective combatants are light infantry.

- Tanks have a limited utility for the counter-guerrilla force, but can serve as an effective reserve on the right terrain. Armored combat vehicles and helicopters can play an important role in mobility and fire support. Mechanized forces usually fight effectively only when dismounted and when using their carriers for support or as a maneuver reserve.

In summary, whatever else these lessons may show, the most fundamental of them is that no army, however sophisticated, well trained, materially rich, numerically overwhelming and ruthless, can succeed on the battlefield if it is not psychologically fit and motivated for the fight. The force; however destitute in material advantages and numbers, relying on the moral qualities of a strong faith, stubborn determination, individualism and unending patience will always be the winner. These may not be the optimum qualities always found in the armies of western democracies fighting in foreign soil [Ref.19].

F. THE RUSSIAN-CHECHEN WAR

In December 1994 Russian authorities made their first attempt to crush Chechen separatism militarily. However, after two years of bloody combat the Russian Army was forced to withdraw from the Chechen Republic. The obstinacy of the Russian authorities that had decided on a policy of victory in Chechnya resulted in the deaths of at least 30,000 Chechens and 5,000 Russian soldiers. This war, which caused an estimated $5.5 billion US dollars in economic damage, was part of the cause of Russia's national economic crisis in 1998, when the Russian government proved unable to service its huge debts [Ref.16].

Several months of Russian attacks shifted the balance of power in Chechnya and changed war analysts' perspective of urban warfare. The tactics employed by both sides are forcing experts to take another look at the concept of urban warfare.

On the other hand, the Chechen rebels were quick to adapt to changing strategic, as well as tactical, situations. Recognizing the value of information operations, they

employed advanced commercial communications systems that actually exceeded the quality of the military gear issued to the Russian army.

The Russians were not careful about decoded talking on their communication systems. Almost all Chechens speak Russian, so they were able to listen to Russian messages and transmit misleading instructions that redirected Russian artillery and aircraft missions. Russian interior troops in particular were careless about unsecured radio communications. In contrast, the Chechens had a native form of secured communications. While virtually every Chechen speaks Russian, almost no Russians speak Chechen. By adding a few simple code phrases, the rebels could communicate in the open without fear of interception. In addition, most of the Chechen radio operators recognized each other by voice.

The Russians basically saw the Chechens as guerillas or bandits who could not stay and fight. This contrasted with the reality of a hardened, dedicated ethnic group possessing of a history of fierce fighting for their longtime homeland against superior powers. Once the Russians entered Grozny, they quickly realized that their adversary was much more determined than expected [Ref.16].

1. Lessons learned

This conflict also presented some of the most important lessons in terms of asymmetric war. These lessons are:

- In times of asymmetric wars, the side, lacking the stronger vast manpower resources, should withdraw from indefensible positions to rugged territory that more easily negates many of the enemy's strengths.
- It is a terrible mistake for both sides to fail to understand the cultural aspects of the conflict
- The lack of small, skilled infantry units at company levels and below plagues an anti-guerrilla force throughout the operation. Coupled with this is the presence of a "strategic corporal" in an anti-guerrilla force that provides experience and leadership for a small unit in combat. Lacking

noncommissioned officers in a highly skilled infantry can become costly in an urban environment. In some cases airborne, naval infantry and Special Forces interior units, who have more cohesive structures, perform better in urban or rural terrain against guerrillas.

- Guerrilla forces must be careful about their radio conversations. Secured communication plays a very important role in asymmetric war environment.
- Targeted use of smaller, more effective infantry units by anti-guerrilla forces will always have its pay-off. These forces are more cohesive. They would engage guerrilla forces to draw them into combat; at which point the anti-guerrilla forces can call in air strikes or artillery on the rebel positions. The specialized forces then would eliminate any surviving resistance.
- Two of the weapon systems used in this conflict proved to be "absolutely devastating" inside the urban environment. One is the RPG, which many Chechens now refer to as their national weapon. The second-most useful guerrilla weapon is the sniper. The Chechens used independent sniper scout teams deployed in small numbers in buildings throughout the city.
- Good cultural intelligence has also grown in importance. This intelligence is valuable both for dealing with adversaries and for understanding noncombatants in the conflict area.

G. TURKEY STRUGGLES WITH TERRORISM

Since its establishment in 1923, Turkey has maintained a vigilant eye on potential threats from across its borders with Bulgaria, Greece, Syria, Iraq, Iran and the former Soviet Union. Today, however, the greatest threat to Turkey's territorial integrity comes from within its borders. This threat is the Kurdish Workers Party (PKK), which is a militant, terrorist, and separatist organization [Ref.10].

The PKK is a terrorist organization that has adopted the Marxist-Leninist ideology and plays the role of the religious fundamentalists. This group depends upon circumstances to weaken and divide Turkey with actions based on violence while attempting to establish a Marxist Kurdish State in Turkey. The PKK terrorist organization uses the method of "Revolutionist Violence" to reach its goal and tries to subdue the people in southeastern Anatolia by terror and pressure. This primitive, savage and bloody terrorist organization, with the claim of protecting the rights of the so-called Kurdish people, has brutally murdered over 30,000 people, including an estimate of 4,000 Turkish soldiers [Ref.9].

The PKK terrorist organization is able to continue its existence through the shelter and monetary support provided by outside countries. Furthermore, the PKK terrorist group organizes aid campaigns through its auxiliary organizations in Europe to increase its financial resources. The PKK conducts this by forcibly collecting funds from the citizens of the Republic of Turkey living in those countries and by engaging in drug smuggling. The money obtained from drug smuggling is a very large share in the total revenues of the organization. The organization is able to purchase weapons, ammunition and equipment, due to the considerable profits obtained from this illegal trade [Ref.9].

The PKK terrorist organization commits all kinds of organized crimes besides its terrorist activities. The organization increasingly resorts to drug smuggling followed by weapons and human smuggling and money laundering activities in order to finance its endeavors [Ref.9].

The terrorist identity of the PKK has been accepted by the main Western countries; led by the United States, United Kingdom, France and Germany. Furthermore, France and Germany banned any activities associated with the PKK and its auxiliary branches in 1993. Upon the prohibitions and restrictions brought by many Western European countries, the organization has shifted these types of operations to Eastern Europe and the Balkans while continuing many of its operations in the Western European countries under the guise of "cultural activity." With this connection, the PKK terrorist organization continues to form many associations, which it claims, have humanitarian purposes [Ref.9].

Turkey has obtained significant successes in recent periods against the terrorist organization, which it has been fighting for over 15 years. In addition, Turkey has taken concrete steps on discontinuing any support provided to the organization from abroad. Due to this effort, Syria, with the Adana Agreement signed on October 20, 1998, has for the first time accepted the PKK to be a terrorist organization agreeing to close down the auxiliary branches of the organization in Syria. Moreover, the founder and chief of the terrorist organization, who was secretly sent to Kenya to hide, was captured on February 16, 1999 and brought to Turkey [Ref.9]. While still ongoing, the tide has turned against the PKK who are increasingly resorting to political means.

1. Lessons learned

This conflict presents some instructive and contemporary aspects of guerrilla warfare. The role of Turkish Army in assisting internal forces to combat the PKK is directly related to operations other than war (OOTW) [Ref.10]. Thus, an examination of the use of military forces during this conflict reflects some of the challenges that may face all conventional forces in future OOTW missions:

- Violent PKK raids on small villages underscored the severity of the internal security threat. Police forces in urban areas, supplemented by gendarmerie operating in the vast mountainous expanse of Southeastern Turkey, were unable to provide adequate security to citizens living in remote settlements. Therefore, the government instituted a Temporary Village Protection (GKK) system in 1985, initiating the formation of a civilian militia under the Interior Minister.

- Under the provisions of the GKK system, village guards are recruited from Kurdish tribes loyal to the state. These paid volunteers, who do not wear distinctive uniforms, receive small arms and minimal training from the government and represent a deterrent state security force in small villages.

- Turkish forces finally realized the importance of intelligence, especially from local people, after having suffered so many military and civilian casualties. Special Forces personnel have been trained as local citizens

and sent out to live in those villages gathering intelligence about activities, locations, and logistics networks of terrorists. This application has served two important purposes. First, timely and precise intelligence have been gathered and second strong relationships have been formed between the government and citizens living in those areas.

- Tanks, Armored Combat Vehicles, and Helicopters have been effectively used against terrorists, especially in suitable terrains. Especially, firepower, sensor ranges, maneuver capabilities, and thermal systems of these assets provide big advantages to conventional forces against guerrillas. Utilizing these assets in full coordination with ground troops and as a fire support for infantry missions increases this effect.

- This conflict also emphasized the significance of night combat and combat under poor weather conditions. In most cases the terrorists prefer to engage with Turkish forces under these conditions. Thermal sight and night sight systems have been widely used.

- Mine warfare was the most challenging issue, causing many problems for Turkish forces. Terrorists have effectively used mine warfare against friendly vehicles and infantry units. Mine locating and clearing is one of the most crucial concerns, awaiting the units encountering these kinds of conflicts.

- Most of the battles that have occurred between the two sides show a striking contrast to the detailed and precise information given in map problems. The situation, particularly in guerrilla warfare, is almost invariably covered in obscurity. Occasionally the information at hand may be false or misleading; therefore, the leader has to recognize the lack of reliable information and prepare for prompt and decisive action.

- Much of the ability of a leader to avoid friendly casualties and to increase the number of enemy casualties is determined prior to the skirmish. So, the responsibilities have to be clearly stated beforehand and given to each

member of the unit. Doubts or questions in the minds of the soldiers should be absent when the battle commences.

- Trade-offs always exists between mobility and firepower of guerrilla forces. Since the mobility of the force is more critical in most cases guerrilla forces should be equipped with light weapons, such as rifles- including sniper rifles- machine guns, RPGs and 60-mm. mortars. Mine detectors and medical gear should be with the platoon. As a standard rule, a guerrilla unit should include between 17 and 23 guerrillas.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. DETAILED MANA DESCRIPTION

A. OVERVIEW

This Appendix explains the details of the MANA model. It gives some key concepts and the details of the parameters in the model.

B. KEY CONCEPTS

1. Squads

A squad is a group of agents of any size. These agents share the same properties, and can either switch between states individually or as a group. A state is a set of parameters values that determine the automaton's current behavior [Ref.1].

Apart from the behavioral and capability parameters, entities in the same squad also share a situational awareness map of enemy contacts, and waypoints that may be used to guide the entities around the battlefield [Ref.1].

2. Battlefield

The default Battlefield for MANA is a 200 x 200 grid of cells, each of which can be occupied by a single entity only. No agents can wander off the battlefield boundaries [Ref.1]. However, not all cells need to be equivalent. Currently, there are three different types of cell: Plain, Easy Going, and Obstacle.

Plain terrain has no special properties. Easy Going terrain represents roads, or other areas that are particularly attractive to move along. Yellow coloring corresponds to these areas. In this version of MANA, an entity in an Easy Going cell experiences no difference when occupying it. However, entities can have their personalities set towards Easy Going terrain. Thus, a "convoy" can be made to stay close to a "road". Gray coloring represents obstacle terrain. No entity may occupy an Obstacle cell. However, in this version, entities can see through Obstacle cells.

Only the correct shade of gray and yellow cause an entity to react to it. However, the main view of the battlefield may be colored to give the impression of open terrain or cover. Such backgrounds have no effect on the entities in the model, and are purely for decorative purpose. Even so, using such an ornamental background helps the user illustrate that the values chosen for the "Stealth" or "Sensor" parameters for a particular scenario represent operating entities in a wooded area with plenty of cover, for example. A background, which appears as woods, helps emphasize this. Despite any decorative visual distractions, Viewing the map as if appears to the entities is easily done by clicking on the View menu at the top of the screen and selecting Terrain as Seen by Agents [Ref.1]. This is shown in Figure 43.

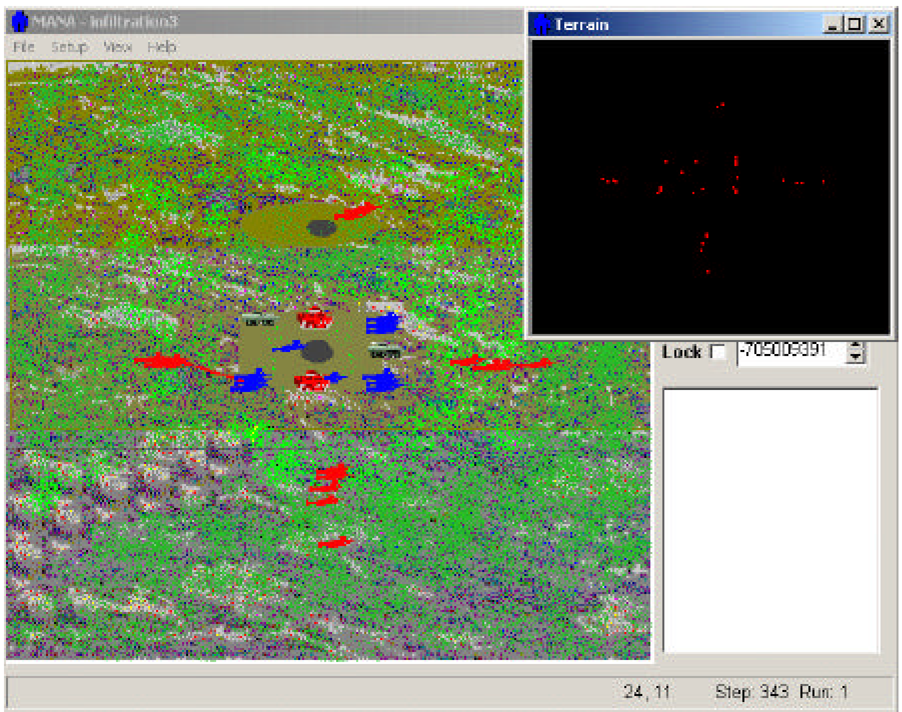

Figure 43. Terrain as seen by agents.

3. Waypoints

Waypoints (shown as flags on the Battlefield) can be set up to guide the movement of entities. An entity's personality settings can be used to attract or repel it from its next waypoint [Ref.1].

4. Situational Awareness

Squads of agents can share"memory" of the enemy squads' locations in the form of a collective picture of sensor information. A squad may share this information with other squads. This is the method of representing inter-squad communication in MANA. A nominated Headquarter (HQ) squad collects all information in its side's collective sensor range in the form a Situation Awareness (SA) map [Ref.1].

A squad's current SA "map" can be shown by selecting Situational Awareness of Squad in the View menu. The arrows at the bottom are used to select the desired view of the squad's SA map [Ref.1]. Figure 44 shows an example:

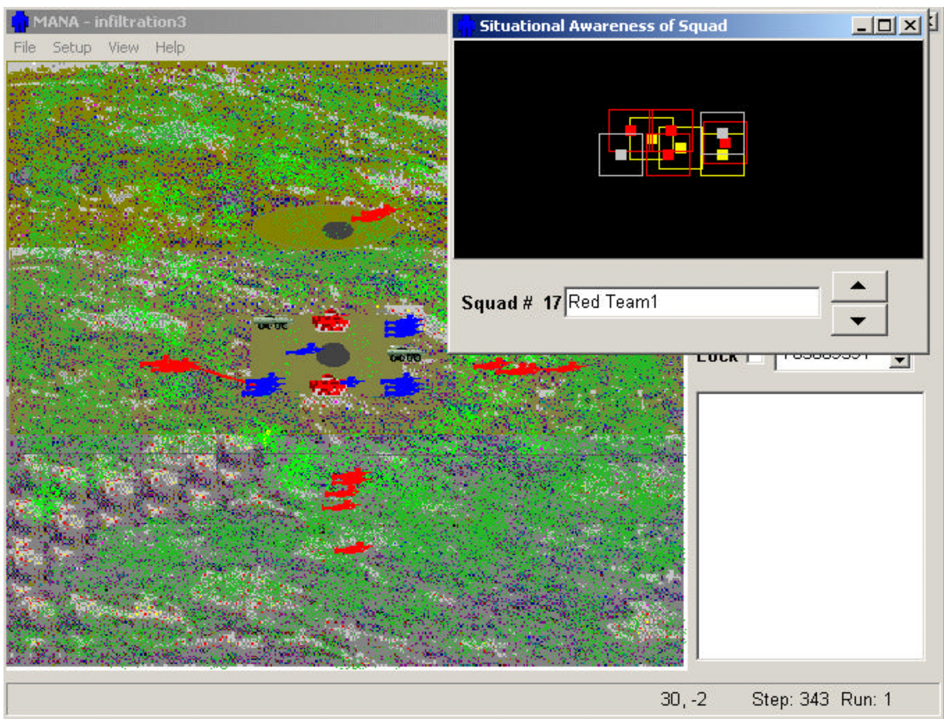

Figure 44. Squad Situational Awareness.

Note that the enemy locations are color-coded; Red represents the highest threat level and fades to yellow, light gray and then disappears with time. The rate of color decay can be set for each squad [Ref.1].

The boxed area on the SA map is the region other agents must be in to react to the information. Thus friendly units can be monitored keep their distance from known

enemy locations by making them retreat whenever they enter the area of the box. A squad's personality parameters can be used to force entities to respond to different threat "levels" using this map and the color-coding scheme [Ref.1].

C. EDITING SQUAD PROPERTIES

The properties of Squads are edited by using the Edit Squad Properties screen under the Setup menu [Ref.1]. There are three separate parts to this screen, each accessed via its own tab:

- **General Properties:** Starting points, goals, number of agents, number of squads and squad home.
- **Personality Properties:** Personality weightings, firepower, stealth and icons.
- **Ranges:** Sensor range, weapon range, movement speed and other constraints.

1. General Squad Properties

General Squad Properties are independent of the Squad's current state [Ref.1]. Figure 45 shows the General Squad Properties screen.

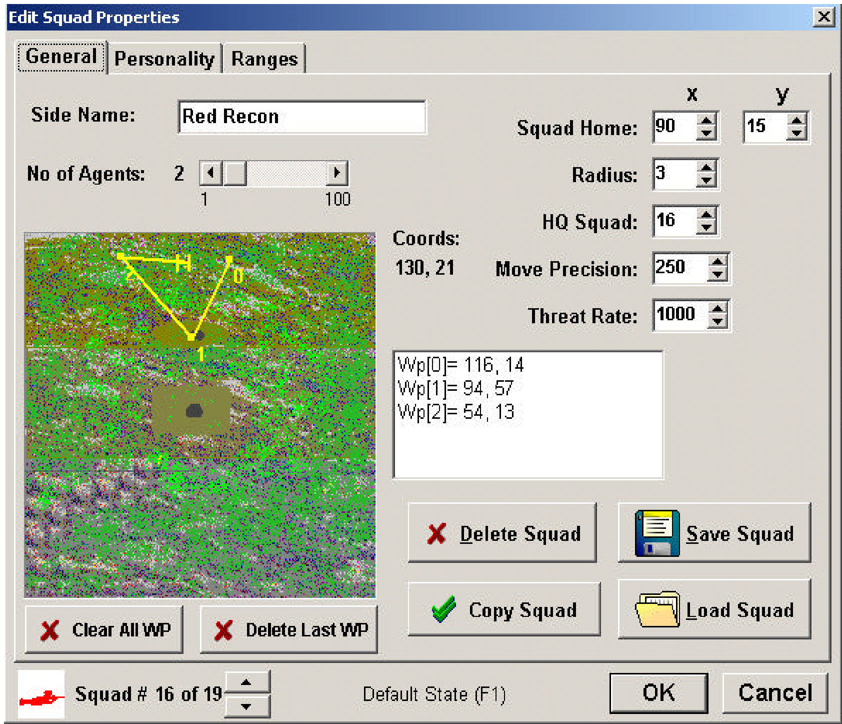

Figure 45. General Squad Properties.

- **Squad #:** This shows the currently selected squad, along with its icon.
- **Side Name:** This provides a name for the current squad.
- **No of Agents:** This allows selection of number of agents in a Squad.
- **Squad Home (x, y):** This is the center of the area within which the squad starts. Members of the Squad are placed at random locations within this area at the start of each run.
- **Radius:** This is the maximum distance from the Home coordinates that the agents are randomly placed on at start up.
- **HQ Squad**: All Squads of the same Allegiance with the same HQ Squad share the same Situational Awareness map. Any and all friendly Squads can be used as an HQ Squad.
- **Move precision:** MANA allows a degree of randomness in the movement of entities. Though the effect of this parameter depends largely on how the other parameters are set, it is particularly useful for preventing Squad

members from moving too rigidly or getting stuck behind obstacles. Additionally, it can cause Squads to become dispersed and to choose a winner (tie breaker) when deciding where to move next when two or more moves are equivalent. A small value makes the entities move more rigidly, whereas a large value increases the randomness.

- **Threat rate:** This is the number of time steps that must pass for a threat on the Situational Awareness map to decay one level. For example, if the threat rate is 40, it will take 40 time steps for a Level 3 threat to turn to a Level 2 threat, and another 40 steps for a Level 2 threat to turn to a Level 1 threat, and another 40 steps to be forgotten altogether. This allows the user to manipulate the entities into reacting differently to older information. Once the units are "forgotten", they disappear from the SA map.
- **Waypoints:** Waypoints are entered straight onto the Map picture, in reverse order, by clicking the left mouse button [Ref.1].

2. Personality Squad Properties

These parameters describe the basic personalities of the entities and allow specifying changes based on circumstances (i.e., how the personalities switch between states) [Ref.1]. Figure 46 shows the Squad Personality screen:

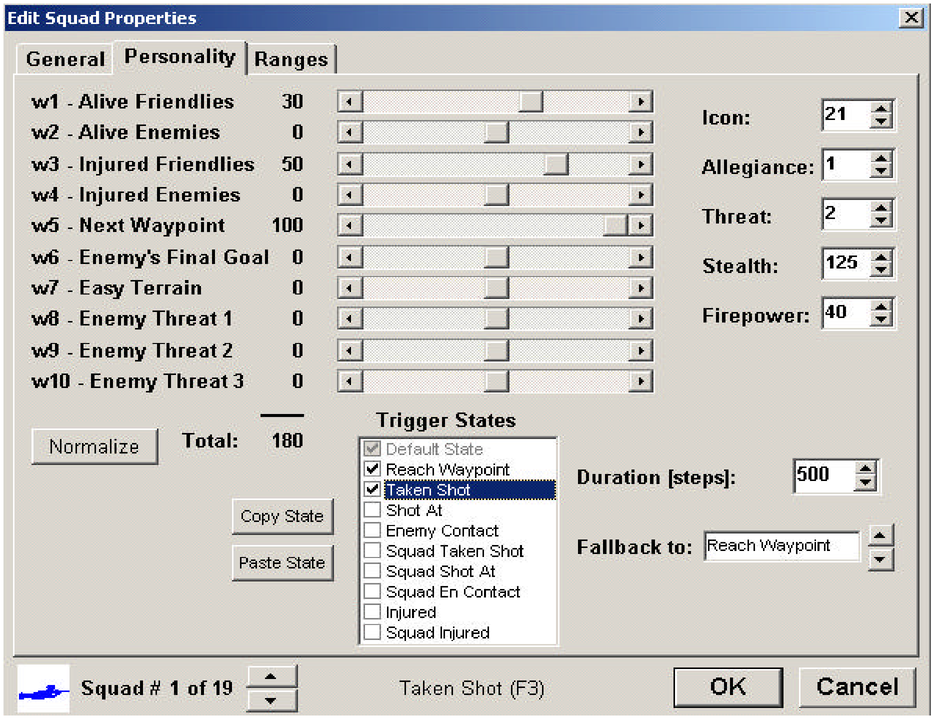

Figure 46. The Squad Personality Tab.

- **Icon:** This selects which icon (displayed in the bottom left-hand corner) is to be used for the Squad in its present state.

- **Allegiance:** This selects which "side" a squad is on in the selected state. By convention Blue Force is Allegiance 1, Red Force is Allegiance 2 and Allegiance 0 is neutral.

- **Threat:** This parameter allows the user some ability to cause the entities to react differently to different types of units recorded on the Situational Awareness map. Units appear on the SA map as red if Level 3, yellow if Level 2, and gray if Level 1. The level of threat falls from 3 to 2, 2 to 1, etc., after a determined period of time. This also allows the user to manipulate the entities to react differently to older information, as well as to different threat types.

- **Stealth:** The Stealth parameter is intended to represent how difficult it is to see an entity once it is within an enemy entity's Sensor range. It represents the degree of cover an entity is assumed to be able to use. An

entity with a Stealth factor of 100 is impossible to see and, therefore, impossible to kill while a Stealth value of 0 makes no difference to visibility or kill probability.

- **Firepower:** This is the probability of killing an enemy entity with a single shot (SSKP).

- **w1 - w10:** These are the weightings entities use to determine their next move. The higher the weighting is, the greater the attraction. Likewise, the more negative the weighting is, the greater the repulsion. The weighting factors affect the way entities react to other entities, Waypoints, terrain and information on the Situational Awareness map. Table 9 lists and describes the weighting factors.

Personality Element	Description	Controls propensity to move toward/away from
w1	Alive Friends	Agents of same Allegiance
w2	Alive Enemies	Agents of enemy Allegiance
w3	Injured Friends	Injured agents of same Allegiance
w4	Injured Enemies	Injured agents of enemy Allegiance
w5	Next Waypoint	The next waypoint agent's squad has been assigned
w6	Enemy's Final Goal	The enemy's final goal (waypoint)
w7	Easy Terrain	Areas deemed Easy Going e.g. yellow road.
w8	Enemy Threat 1	Enemies in SA map which are of Threat Level 1
w9	Enemy Threat 2	Enemies in SA map which are of Threat Level 2
w10	Enemy Threat 3	Enemies in SA map which are of Threat Level 3

Table 9. Description of Weighting Factors

- **Trigger States**: This determines which state an entity is currently in. The state, in turn, determines the entity's appearance (icon), Allegiance, Threat Level, Stealth, Firepower, and Personality Weighting values as well as the values used on the Ranges Squad Properties screen.

- **Duration:** This is the number of time steps a squad will stay in a state once that state has been triggered.

- **Fallback State:** Once the Duration time has expired, the entity automatically falls back to the Default state, unless otherwise specified. The Fallback State box allows the user to choose which state the entity falls back [Ref.1].

3. The Ranges Squad Properties

As with the Personality Squad Properties, the Ranges Squad Properties are dependent on the current entity/Squad state [Ref.1]. Figure 47 shows the Ranges Squad Screen:

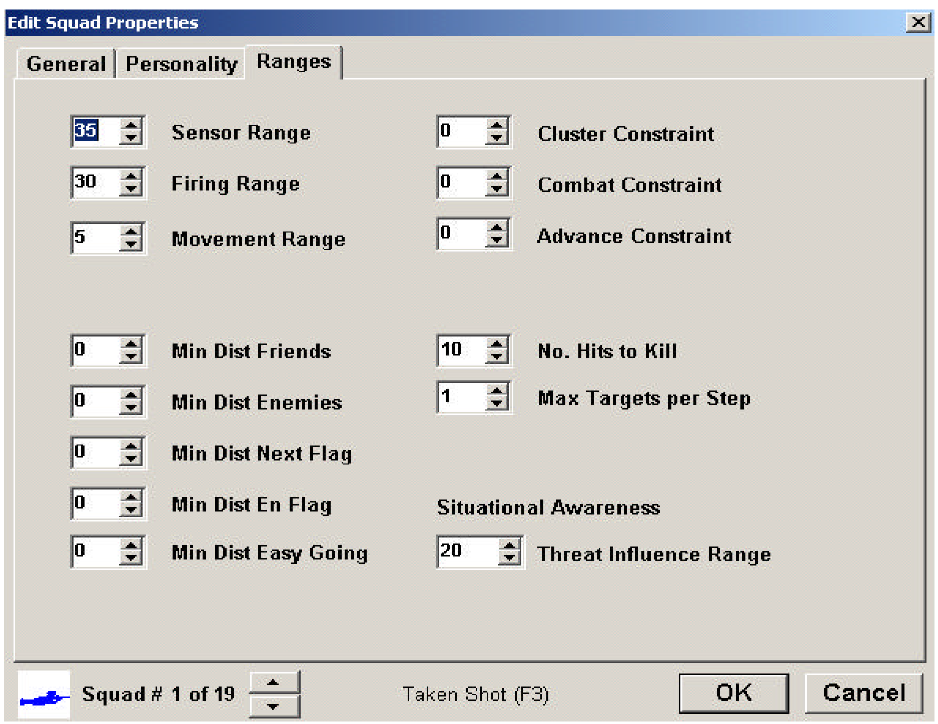

Figure 47. The Ranges Squad Properties Tab.

- **Sensor Range:** An entity can identify any other entity or terrain with this number of cells.

- **Firing Range:** An entity can shoot at any other entity within this number of cells.
- **Movement Speed:** An entity can move this number of cells in a single time step (0 to 100 cells).
- **Min Distance to Friends:** This parameter limits the distance to which friendly entities approach each other. The value entered is the minimum distance, in terms of number of cells that the entities try to maintain.
- **Min Distance from Enemies:** As with Min Distance to Friends, this parameter is the minimum distance (measured in number of cells) that an entity tries to keep between it and the enemy.
- **Min Distance from Next Flag:** This is similar to the other Min Distance parameters, except that this is the minimum distance to the next Waypoint.
- **Min Distance from Enemy Flag:** This is similar to the other Min Distance parameters, except that this is the minimum distance to the Enemy's Final Waypoint, referred to as the Enemy Flag.
- **Min Distance from Easy Going:** This is the minimum distance that an agent keeps from Easy Going terrain.
- **Cluster Constraint:** The cluster constraint is designed to prevent agents from clustering in groups larger than a specified size, as determined by the value of this parameter. It operates on the attraction towards friends. Once the friend's penalty component, which is defined later in this chapter, of a move is calculated, it is only counted if the number of friends within sensor range is greater than the cluster constraint. Otherwise, the penalty is set to zero, thus ignoring friends if in a group larger than the cluster constraint size. If the cluster constraint is zero or if the personality component towards friends is negative (repelled from friends) then the cluster constraint is ignored.

- **Combat Constraint:** The combat constraint permits advancing toward the enemy only if a certain numerical advantage is met, i.e., outnumber the enemy by a number of agents. The combat constraint modifies the penalty component for moving toward enemies and injured enemies. The numerical advantage is calculated as number friends – number enemy (in sensor range). If the difference is greater than the combat constraint, then the move toward enemy penalty is counted, otherwise it is negated, i.e., an attraction toward enemy becomes repulsion. If the combat constraint is zero or if the personality weight towards the enemy is negative (repelled from the enemy), then the combat constraint is ignored.
- **Advance Constraint:** The advance constraint stops the penalty component from moving toward the next goal (waypoint). This constraint is intended to prevent a Squad advancing towards its next goal without sufficient support. If the number of friends within an agent's sensor range is fewer than the advance constraint, then the penalty toward the next goal is negated i.e., repelled from next goal, rather than attracted.
- **Number of Hits to Kill:** This identifies the number of hits an entity can sustain before it is deemed to be killed. If the number of hits is greater than one then the first hit puts the entity into the Injured State. Subsequent hits have no further effect until the number of hits to kill is reached.
- **Max Targets per Step:** This determines the number of targets within both sensor and firing range that can be shot at in a single time step.
- **Threat Influence Range:** As discussed in the section on Situational Awareness, whenever an enemy entity is marked on the Situational Awareness map, a boxed area appears around it. Within this boxed region, agents are able to react to the recorded enemy location. The reason for restricting the effect of SA information to a designated area around the enemy target is to allow entities to keep their distance from known enemy positions, without necessarily running away from them. That is, they only run away if within the Influence Range area. Likewise, if the entity's

personality is such that it chases enemy entities, the Influence Area is supposed to represent how near the entity needs to be before it will bother to chase. The Threat Influence Range is simply the number of cells this area is wide [Ref.1].

APPENDIX C. CLUSTER ANALYSIS REPORTS

A. OVERVIEW

This Appendix presents the results of Cluster Analysis completed in Clementine. As previously stated two data sets exist in our analysis: the 22-factor data set and the 6-factor data set. The purpose of the Cluster Analysis is to compare our MOEs to explore the patterns associated with them. In this analysis, the MOEs are partitioned into three categories. For both MOEs, the values from 0 to 0.25 are assigned a value of 0 (low), the values from 0.25 to 0.75 are assigned a value of 1 (medium), and the values bigger than 0.75 are assigned a value of 2 (high).

B. THE 22-FACTOR DATA SET

In the clustering report below five clusters exist. In Cluster-1, 24 percent of the data, Red inflicts medium Blue casualties while suffering high casualties approximately 50 percent of the time. In Cluster-2, 2.5 percent of the data, both sides suffer low casualties. In Cluster-3, 1 percent of the data, Blue takes low casualties while inflicting medium Red loses. In Cluster-4, 1 percent of the data, Red inflicts high casualties on Blue while taking medium or high casualties. Finally, in Cluster-5, 71.5 percent of the data, Blue lose few agents while Red lose many agents.

```
Number of inputs = 6
Number of records = 51300
Initial number of clusters = 5
Final number of clusters = 5

cluster 1: 12347 examples
  Blue Prop :
    1 -> 1.0
  Red Prop :
    0 -> 0.14708
    1 -> 0.319754
    2 -> 0.533165
-----------------------------
cluster 2: 1253 examples
  Blue Prop :
    0 -> 1.0
```

```
  Red Prop :
    0 -> 1.0
-----------------------------
cluster 3: 505 examples
  Blue Prop :
    0 -> 1.0
  Red Prop :
    1 -> 1.0
-----------------------------
cluster 4: 577 examples
  Blue Prop :
    2 -> 1.0
  Red Prop :
    0 -> 0.017331
    1 -> 0.242634
    2 -> 0.740035
-----------------------------
cluster 5: 36618 examples
  Blue Prop :
    0 -> 1.0
  Red Prop :
    2 -> 1.0
-----------------------------
```

C. THE 6-FACTOR DATA SET

Another Cluster Analysis is performed for the 6- factor data set. Five clusters also exist in this analysis. In Cluster-1, 67 percent of the data, Blue takes low casualties most of the time while causing high Red loses. In Cluster-2, 22 percent of the data, Red inflicts medium Blue casualties while suffering low casualties. In Cluster-3, only 4 observations, Red takes low casualties while inflicting high Blue casualties. In Cluster-4, 0.8 percent of the data, Red suffers low casualties while Blue takes high casualties. Finally, in Cluster-5, 10 percent of the data, both sides take low casualties.

```
Number of inputs = 6
Number of records = 24300
Initial number of clusters = 5
Final number of clusters = 5

cluster 1: 16182 examples
  Blue Prop :
    0 -> 0.9285
    1 -> 0.068224
    2 -> 0.003275
```

```
  Red Prop :
    2 -> 1.0
-----------------------------
cluster 2: 5267 examples
  Blue Prop :
    1 -> 1.0
  Red Prop :
    0 -> 0.997341
    1 -> 0.002658
-----------------------------
cluster 3: 4 examples
  Blue Prop :
    2 -> 1.0
  Red Prop :
    1 -> 1.0
-----------------------------
cluster 4: 197 examples
  Blue Prop :
    2 -> 1.0
  Red Prop :
    0 -> 1.0
-----------------------------
cluster 5: 2650 examples
  Blue Prop :
    0 -> 1.0
  Red Prop :
    0 -> 1.0
-----------------------------
```

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. NEURAL NETWORK REPORTS

A. OVERVIEW

This Appendix gives the results of the Neural Network models developed in Clementine. Two data sets exist: the 22-factor data set and the 6-factor data set.

B. THE 22-FACTOR DATA SET

A neural network is created for the 22-factor data set. The data set is split into training and test sets. The training set includes 10 percent of the data. The test set contains another 10 percent of the data.

1. Red Killed

The first model is developed and tested for MOE1, the proportion of Red killed. The report below illustrates the size of the different layers of the network, its predicted accuracy on the training set and the relative importance of each variable in predicting the MOE in the model.

a. Training

```
Neural Network "Red Prop" architecture

Input Layer     : 22 neurons
Hidden Layer #1 : 8 neurons
Output Layer    : 1 neurons
Predicted Accuracy :  96.72%

Relative Importance of Inputs
 Red Stealth    : 0.49745
 Red Movement   : 0.19273
 Red w2         : 0.16879
 Vehicles Firing: 0.16734
 Red NumAgents  : 0.16699
 Red SSKP       : 0.15990
 Red w1         : 0.14304
 Red w10        : 0.13651
 Red Combat     : 0.13354
 Blue   Stealth : 0.13155
 Red w3         : 0.13120
```

```
Blue   Sensor  : 0.12004
Recon Firing   : 0.11812
Recon SSKP     : 0.11138
Vehicles Sensor: 0.10753
Recon Max Trgts: 0.10676
Red w8         : 0.10652
Recon Sensor   : 0.10158
Recon Stealth  : 0.10064
Blue   Firing  : 0.09517
Red Cluster    : 0.09507
Blue     SSKP  : 0.08039
```

b. Prediction on the Test Set

After developing the model, the test set is used to see its predictive power. In this report, especially, the mean absolute error is of interest, which is 0.033.

```
Results for output field Red Prop
    Error analysis for $N-Red Prop
        Minimum Error       :        -0.43703
        Maximum Error       :         0.28549
        Mean Error          :        -0.0097315
        Mean Absolute Error :         0.033438
        Standard Deviation  :         0.052032
        Linear Correlation  :         0.97615
        Occurences          :         10234
```

2. Blue Killed

Another network model is developed and tested for MOE2, the proportion of Blue killed. The report below illustrates the size of the different layers of the network, its predicted accuracy on the training set and the relative importance of each variable in predicting the MOE in the model.

a. Training

```
Neural Network "Blue Prop" architecture

Input Layer     : 22 neurons
Hidden Layer #1 : 8 neurons
Output Layer    : 1 neurons

Predicted Accuracy :  95.23%
```

```
Relative Importance of Inputs
 Red Stealth    : 0.30289
 Red NumAgents  : 0.22495
 Red SSKP       : 0.18002
 Recon Firing   : 0.14796
 Recon Stealth  : 0.12291
 Red Combat     : 0.11536
 Recon SSKP     : 0.11012
 Red Movement   : 0.10599
 Blue     SSKP  : 0.10463
 Blue   Stealth : 0.10459
 Recon Sensor   : 0.10362
 Recon Max Trgts: 0.09819
 Vehicles Firing: 0.09808
 Blue   Firing  : 0.09374
 Red w1         : 0.08430
 Vehicles Sensor: 0.07851
 Red w10        : 0.07825
 Red Cluster    : 0.07091
 Red w2         : 0.06928
 Red w3         : 0.06421
 Blue   Sensor  : 0.05188
 Red w8         : 0.05113
```

b. Prediction on the Test Set

After developing the model, the test set is used to see its predictive power. In this report, especially, the mean absolute error is of interest, which is 0.048.

```
Results for output field Blue Prop
    Error analysis for $N-Blue Prop
        Minimum Error       :        -0.55604
        Maximum Error       :         0.78042
        Mean Error          :         0.0041290
        Mean Absolute Error :         0.048269
        Standard Deviation  :         0.094468
        Linear Correlation  :         0.90514
        Occurrences         :         10234
```

C. THE 6-FACTOR DATA SET

A neural network is created for the 6-factor data set. The data set is split into training and test sets. The training set includes 80 percent of the data. The test set contains another 20 percent of the data.

1. Red Killed

The first model is developed and tested for MOE1, the proportion of Red killed. The report below illustrates the size of the different layers of the network, its predicted accuracy on the training set and the relative importance of each variable in predicting the MOE in the model.

a. Training

```
Neural Network "Red Prop" architecture

Input Layer     : 6 neurons
Hidden Layer #1 : 8 neurons
Output Layer    : 1 neurons

Predicted Accuracy :  98.72%

Relative Importance of Inputs
Red Stealth          : 0.92148
Recon Stealth        : 0.09658
 Red Num             : 0.08013
Red SSKP             : 0.00691
Recon Sensor         : 0.00188
Recon Firing         : 0.00185
```

b. Prediction on the Test set

After developing the model, the test set is used to see its predictive power. In this report, especially, the mean absolute error is of interest, which is 0.013.

```
Results for output field Red Prop
    Error analysis for $N-Red Prop
        Minimum Error       :        -0.23586
        Maximum Error       :         0.064019
        Mean Error          :        -0.0020265
        Mean Absolute Error :         0.012959
        Standard Deviation  :         0.026350
        Linear Correlation  :         0.99816
        Occurrences          :         4819
```

1. Blue Killed

Another network model is developed and tested for MOE2, the proportion of Blue killed. The report below illustrates the size of the different layers of the network, its predicted accuracy on the training set and the relative importance of each variable in predicting the MOE in the model.

a. Training

```
Neural Network "Blue Prop" architecture

Input Layer     : 6 neurons
Hidden Layer #1 : 8 neurons
Output Layer    : 1 neurons

Predicted Accuracy :  95.97%

Relative Importance of Inputs
Red Stealth          : 0.51905
Red SSKP             : 0.28759
 Red Num             : 0.18888
Recon Stealth        : 0.06832
Recon Firing         : 0.05855
Recon Sensor         : 0.05846
```

b. Prediction on the Test set

After developing the model, the test set is used to see its predictive power. In this report, especially, the mean absolute error is of interest, which is 0.104.

```
Results for output field Blue Prop
    Error analysis for $N-Blue Prop
        Minimum Error       :        -0.65080
        Maximum Error       :         0.73333
        Mean Error          :        -0.023202
        Mean Absolute Error :         0.10456
        Standard Deviation  :         0.19476
        Linear Correlation  :         0.87471
        Occurrences          :         4819
```

APPENDIX E. LHD 22 FACTORS

A. OVERVIEW

This Appendix shows and explains the 22 parameters chosen in MANA for analysis purposes. It also provides the ranges of each parameter.

MANA Scenario File (Version 0.97a)

February 14, 2002

A. Red Team 1,2 (Squad 17,18) / No. of agents – No. of agents belonging to a squad

B. Red Team 1,2 (Squad 17,18) / enemy contact / w1 – Controls propensity to move towards agents of same allegiance

C. Red Team 1,2 (Squad 17,18) / enemy contact / w2 – Controls propensity to move towards agents of enemy allegiance

D. Red Team 1,2 (Squad 17,18) / enemy contact / w3 – Controls propensity to move towards injured agents of same allegiance

E. Red Team 1,2 (Squad 17,18) / enemy contact / w8 – Controls propensity to move towards enemies in SA map which are of threat level 1

F. Red Team 1,2 (Squad 17,18) / enemy contact / w10 – Controls propensity to move towards enemies in SA map which are of threat level 3

G. Blue Squad 1-11(1 through 11) / taken shot / stealth – Probability I will not be seen

H. Blue Squad 1-11 / taken shot / fire power – Single shot kill probability

I. Red Recon (Squad16) / shot at / stealth – Probability I will not be seen

J. Red Recon (Squad16) / shot at / fire power – Single shot kill probability

K. Red Team 1,2 (Squad 17,18) / enemy contact / stealth – Probability I will not be seen

L. Red Team 1,2 (Squad 17,18) / enemy contact / fire power – Single shot kill probability

M. Blue Squad 1-11 / taken shot / sensor range – No. of grids I can see other agents

N. Blue Squad 1-11 / taken shot / firing range – No. of grids I can shoot

O. Blue Squad 12-15 / taken shot / sensor range – No. of grids I can see other agents

P. Blue Squad 12-15 / taken shot / firing range – No. of grids I can shoot

Q. Red Recon (Squad16) / shot at / sensor range – No. of grids I can see other agents

R. Red Recon (Squad16) / shot at / firing range – No. of grids I can shoot

S. Red Recon (Squad16) / shot at / Max tgts per Step – No. of enemy I can shoot per step

T. Red Team 1,2 (Squad 17,18) / enemy contact / combat constraint – If (friends - enemies) <= combat, then reverse positive w2

U. Red Team 1,2 (Squad 17,18) / enemy contact / movement range – No. of grids I can move per time-step

V. Red Team 1,2 (Squad 17,18) / enemy contact / cluster constraint – If (friends) > cluster, then ignore positive w1

Notes:

Factor A will have settings of 4 to 35 with 129 levels evenly spread out over these values

Factors B-F will have settings of -64 to 64 in increments of 1 that will correspond to 129 levels

Factors G-R will have settings of 1-129 in increments of 1 that will correspond to 129 levels

Factor S will have settings of 1 to 15 with 129 levels evenly spread out over these values

Factor T will have settings of -10 to 9 with 129 levels evenly spread out over these values

Factor U will have settings of 20 to 148 in increments of 1 that will correspond to 129 levels

Factor V will have settings of 0 to 35 with 129 levels evenly spread out over these values

APPENDIX F. FRACTIONAL DESIGN 6 FACTORS

A. OVERVIEW

This Appendix presents the six factors used in further analysis of the data and their levels.

MANA Scenario File (Version 0.97a)
April 21, 2002

A. Red Team 1,2 (Squad 17,18) / No. of agents – No. of agents belonging to a squad

B. Red Recon (Squad16) / shot at / stealth – Probability I will not be seen

C. Red Team 1,2 (Squad 17,18) / enemy contact / stealth – Probability I will not be seen

D. Red Team 1,2 (Squad 17,18) / enemy contact / fire power - SSKP

E. Red Recon (Squad16) / shot at / sensor range – No. of grids I can see other agents

F. Red Recon (Squad16) / shot at / firing range – No. of grids I can shoot

Notes:
Factor A will have settings of 4, 20, and 35 (low, medium, high)

Factors B-F will have settings of 0, 65, and 129 (low, medium, high)

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX G. NETICA RANGES OF VARIABLES

A. OVERVIEW

This Appendix gives the ranges of variables used in Bayesian Network models.

	low		medium		high	
Red Killed	0	0.25	0.25	0.75	0.75	1
Blue Killed	0	0.25	0.25	0.75	0.75	1
Red NumAgents	4	14	15	24	25	35
Red w1	-64	-22	-21	21	22	64
Red w2	-64	-22	-21	21	22	64
Red w3	-64	-22	-21	21	22	64
Red w8	-64	-22	-21	21	22	64
Red w10	-64	-22	-21	21	22	64
Infantry Stealth	1	43	44	86	87	129
Infantry SSKP	1	43	44	86	87	129
Red Stealth	1	43	44	86	87	129
Red SSKP	1	43	44	86	87	129
Recon Stealth	1	43	44	86	87	129
Recon SSKP	1	43	44	86	87	129
Infantry Sensor	1	43	44	86	87	129
Infantry Firing	1	43	44	86	87	129
Vehicles Sensor	1	43	44	86	87	129
Vehicles Firing	1	43	44	86	87	129
Recon Sensor	1	43	44	86	87	129
Recon Firing	1	43	44	86	87	129
Recon Max Trgts	1	5	6	10	11	15
Red Combat	-10	-3	-2	3	4	9
Red Movement	20	62	63	105	105	148
Red Cluster	0	11	12	23	24	35

Table 10. Bayesian Network Ranges

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

1. Michael K. Lauren and Roger T. Stephen, "Map Aware Non-Uniform Automata: MANA Version 1.0 Users Manual," Defense Technology Agency, New Zealand, June 2001.

2. Gary Horne and Mary Leonardi, "Maneuver Warfare," Marine Corps Combat Development Command, 2001.

3. Michael K. Lauren, "Characterising the Difference Between Complex Adaptive and Conventional Combat Models," Defence Technology Agency, New Zealand, 1999.

4. Michael F. Ling, "Agent Based Modeling of Command and Control Effectiveness," Aeronautical and Maritime Research Laboratory, Defense Science and Technology Organization, Australia, 2001.

5. Ian F. W. Beckett, "Encyclopedia of Guerrilla Warfare," ABC-CLIO Inc., 1999.

6. Andrew Ilachinski, "Irreducible Semi-Autonomous Adaptive Combat (ISAAC): An Artificial Life Approach to Land Warfare," Center for Naval Analysis Research Memorandum CRM 97-61, 1997.

7. Jerome Bracken, Moshe Kress, Richard E. Rosenthal, "Warfare Modeling," Military Operations Research Society, John Wiley & Sons, Inc., 1995.

8. Major Scott R. McMichael, "A Historical Perspective on Light Infantry," U.S. Army Command and General Staff College, Fort Leavenworth, Kansas, September 1987.

9. White Book, "Section 6-Terrorism," Turkish National Defense Ministry, 2000.

10. Major Stephen H. Button, "Turkey Struggles With Kurdish Separatism," Military Review, Dec 1994-Jan/Feb 1995.

11. US Army Publication, "FM100-20, Military Operations in Low Intensity Conflicts," Headquarters Department of the Army, December 1990

12. US Army Publication, "RB100-39, Low Intensity Conflict," US Army Command and General Staff College, Fort Leavenworth, Kansas, December 1983.

13. US Army Publication, "96-3, Small Unit Night Fighter Manual," Center for Army Lessons Learned (CALL), Fort Leavenworth, Kansas, March 1996.

14. US Army Publication, "95-5, Winning in the Jungle," CALL, Fort Leavenworth, Kansas, May 1995.

15. US Army Publication, "FM100-5, Operations," Headquarters Department of the Army, June 1993.

16. US Army Publication, "99-16, Urban Combat Operations," CALL, Fort Leavenworth, Kansas, November 1999.

17. The DTIC Review (DTIC/TR-99-12 Vol.4, No.4), "Urban Warfare," August 1999.

18. The DTIC Review (DTIC/TR-97-3 Vol.3, No.1), "Terrorism, Peacekeeping, and Operations Other Than War," Defense Technical Information Center, March 1997.

19. General (ret) Mohammad Yahya Nawroz and LTC (ret) Lester W. Grau, "The Soviet War in Afghanistan: History and Harbinger of Future War," Foreign Military Studies Office, Fort Leavenworth, Kansas, 1996.

20. Data Analysis Product Division, "S-PLUS 2000 User's Guide," MathSoft, Inc., Seattle, Washington, May 1999.

21. Netica API Programmer's Library, "Reference Manual Version 1.06," Norsys Software Corp., May 2002.

22. Marco Ramoni and Paola Sebastiani, "Bayesian Methods for Intelligent Data Analysis," Knowledge Media Institute, July 1998.

23. Clementine Data Mining System Version 6.0.2, "Clementine Tutorial," SPSS Inc., 2000.

24. Deborah F. Swayne, Dianne Cook, Andres Buja and Duncan Temple Lang, "Ggobi Manual," February 2002.

25. Jerome H. Friedman, "Getting Started with MART in S-Plus," Stanford University, July 2000.

26. Averill M. Law and W. David Kelton, "Simulation Modeling and Analysis," The McGraw-Hill Companies, Inc., 2000.

27. George E. P. Box, William G. Hunter and J. Stuart Hunter, "Statistics for Experimenter," John Wiley and Sons, Inc., 1978.

28. Lawrence C. Hamilton, "Regression with Graphics," Wadsworth, Inc., Belmont, California, 1992.

29. David Hand, Heikki Mannila and Padhraic Smyth, "Principles of Data Mining," The MIT Press, Cambridge, Massachusetts, 2001.

30. Jay L. Devore, "Probability and Statistics for Engineering and the Sciences," Duxbury, Thomson Learning, 2000.

31. Robert A. McLean and Virgil L. Anderson, "Applied Factorial and Fractional Designs," Marcel Dekker Inc., 1984.

32. R. L. Iman, "Computer-Intensive Statistical Methods Simple Random Sampling Stratified Designs," Sandia National Laboratories, Albuquerque, N. Mex., 1980.

33. Gregory D. Wyss and Kelly H. Jorgensen, "A User's Guide to LHS," Sandia National Laboratories, Albuquerque, N, Mex., February 1998.

34. Engineering Statistics Handbook, http://www.itl.nist.gov/div898/handbook.

35. P.A.Bassett, and G.Bishop, "Generalized Additive Models," BioSS, Inc., 1987.

36. SAS/ETS User's Guide, Version 8, Volumes 1 and 2, http://www.sas.com.
37. http://users.erols.com/mwhite28/warstat4.htm.
38. http://www.soft.net.uk/loseby/burmastar/chindits.htm.
39. http://ca.essortment.com/malayanemergenc_rwnt.htm.
40. http://www.gunplot.net/malayaindon/indoconf.htm.
41. Muzaffer Coban, “Predicting Battle Outcomes with Classification Trees,” Naval Postgraduate School, December 2001.
42. Lloyd P. Brown, “Agent Based Simulation as an Exploratory Tool in the Study of Human Dimension,” Naval Postgraduate School, March 2000.
43. Tom Cioppa, “Efficient Nearly Orthogonal and Space-Filling Experimental Designs for High-Dimensional Complex Models,” Naval Postgraduate School, June 2002.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   8725 John J. Kingman Rd., STE 0944
   Ft. Belvoir, Virginia 22060-6218

2. Dudley Knox Library
   Naval Postgraduate School
   411 Dyer Rd.
   Monterey, California 93943-5101

3. Genel Kurmay Baskanligi
   Personel Baskanligi
   Bakanliklar
   Ankara, TURKEY

4. Kara Kuvvetleri Komutanligi
   Personel Daire Baskanligi
   Bakanliklar
   Ankara, TURKEY

5. Kara Kuvvetleri Komutanligi
   Kara Kuvvetleri Kutuphanesi
   Bakanliklar
   Ankara, TURKEY

6. Kara Harp Okulu Komutanligi
   Cumhuriyet Sitesi
   Kara Harp Okulu Kutuphanesi
   Bakanliklar
   Ankara, TURKEY

7. Dr. Alfred Brandstein
   MCCDC (C 45), Studies & Analysis Division
   3300 Russell Road
   Quantico, VA 22134-5001

8. Dr. Gary Horne
   The MITRE Corporation
   2750 Killarney Dr
   Woodbridge, VA 22192

9. Mr. Steve Upton
The MITRE Corporation
2907 W Bay to Bay
Tampa, FL 33629

10. Ms. Mary McDonald
SAIC
1100 N Glebe Rd, Suite 1100
Arlington, VA 22201

11. Professor Tom W. Lucas
Department of Operations Research, code OR/LT
Naval Post Graduate School
Monterey, CA 93943

12. Lieutenant Colonel Saverio Manago
Department of Operations Research
Naval Post Graduate School
Monterey, CA 93943

13. Professor Dan Boger
Graduate School of Business & Public Policy
Naval Post Graduate School
Monterey, CA 93943

14. Professor Susan Sanchez
Department of Operations Research
Naval Post Graduate School
Monterey, CA 93943

15. Lieutenant Colonel Tom Cioppa
TRADOC Analysis Center
Naval Post Graduate School
Monterey, CA 93943

16. Ustegmen Arif Ilker Ipekci
Genel Kurmay Baskanligi
Bakanliklar
Ankara, TURKEY